

Commentary

TRANSPORT

WHEN you and I want a loan, the bank manager's practical interest is related to the security we can offer. When a large corporate undertaking needs money from the Government, Parliament asks for a 'report'. Last December the House of Commons, in connection with the Transport (Borrowing Powers) Bill, asked for a report on the plan for modernizing British Railways and the report of the British Transport Commission has now been published as a White Paper (*Comm. 813*). The publication *British Transport in 1958* costs 2s, covers much of the ground and is well worth the money.

If your experience of the railways is limited to trying to book a sleeper from London to Scotland in August or to joining the Torbay express to London at Exeter and trying to get lunch on a Friday in the summer, you are likely to have a distorted view. It seems extremely popular and very profitable for the British Transport Commission. But, of course, the work of the railways is not restricted to the extremes of popular holiday travel. Nor fortunately is the comfort and catering limited to such circumstances. British Transport is a large, important and complex organization. Progress was certainly needed. Progress is being made and showing signs of acceleration. Progress in electrification, in locomotives and rolling stock, particularly brake-fitted freight tracks, in signalling and train control, in freight handling, marshalling and control. The station at Crewe is being reorganized and reconstructed and I have discovered a first-class hotel at Nantwich. In the foreseeable future we shall all be spared the uncivilized horror of steam trains using tunnels. It is grotesque for the engine crew, purgatory for the passengers and disastrous for the well-being of the tunnel. Is there another country in Europe which permits it?

Naturally the railways draw attention to their problems as well as their triumphs. Comparison with competitors for passenger traffic yields surprises. The number of private cars is, of course, increasing; at the end of 1958 it was almost 300 per cent of the number in 1938. The seating capacity of passenger rolling stock at the end of 1958 was, however, rather less than in 1938 and it is curious that the aggregate seating capacity of buses and coaches has risen little since 1954 when it was about 175 per cent of that in 1938. Fares must be judged in terms of money values and wages; thus *British Transport in 1958* shows in diagrammatic form that the number of miles of travel which could be purchased by an hour's work has risen from 24.5 in 1938 to 42.5 in 1958.

The problem of the 'peak load' is one which haunts many public enterprises, the suppliers of power no less than the suppliers of transport. It is becoming more acute; for instance the percentage of passengers leaving Southern Region London termini in the evening peak hour compared with the rest of the day has risen from 31 per cent in 1938 to 43 per cent in 1958. This, of course, involves a grossly inefficient use of equipment and manpower. There is moreover one feature peculiar to the railways. The season ticket holder is, in general, one who travels at peak hours; he is also travelling at discount rates. The economic anomaly is then that the bulk of the railway peak traffic, which is wasteful since much of the necessary equipment is idle during the rest of the day, is provided at great cost and gives a low return.

The two important developments in freight carrying are the increased number of brake-controlled trains and the improvement of marshalling yards and freight control. Essentially these follow the need and the desire to improve the efficiency of use of the railways. The essential features of a railway system are track, rolling stock and motive power. The present purpose is not to extend but to improve the efficiency of use of each, a policy which in its essentials is obviously sound. Track is being relayed and strengthened, more than 8000 miles in the present five years, automatic signalling and control is being extended. Freight carrying, and just as important, handling and marshalling, are being speeded up. The immediate expansion of diesel traction and electrification is the reply to the question of motive power. All this is costing a great deal of money; the limit of deficit borrowing has been extended from £250 millions to £400 millions. The question is whether the rationalizing is proceeding quickly enough.

The railways have a lot to do. There are signs that they are tackling their problems with energy and vigour. This makes it the more deplorable that their thinking at platform level seems inadequate. It is deplorable that it requires what is effectively a strike by many passengers on the London Underground before the reasons, sensible enough when they are forthcoming, for cancelling trains are given to the customers. The means of handling luggage and mail are primitive. No doubt the coveys of train-spotters are a nuisance in many stations but in a sense the travelling public are all spotters. We might be persuaded to take a more extensive interest in the railways and what they are doing by displays and diagrams. Goodwill is worth much and with the public Gresham's law holds.

BUNSEN, KIRCHHOFF AND THE SPECTROSCOPE (1859)

J. H. S. GREEN

IN spite of a certain traditional rivalry, physicists and chemists have sometimes collaborated explicitly in notable contributions to science. The work of Lord Rayleigh and Sir William Ramsay leading to the discovery of the rare gases is a well known example. On the other hand, the collaboration of the chemist R. W. Bunsen and the physicist G. W. Kirchhoff thirty-five years earlier contributed substantially to both sciences and was an important element in the newly arising science of physical chemistry.

Early Work on the Spectrum

The work of Bunsen and Kirchhoff was the culmination of many scattered observations by numerous workers and these early observations¹ are of interest in themselves as well as constituting the background against which the later work must be seen. The production of colours by refraction and its natural manifestation, the rainbow, had of course been known from ancient times. By the early thirteenth century Robert Grosseteste was writing on the theory of the rainbow and the study of the rainbow, optics and colour was a theme continuing into the beginning of modern science in the work of Descartes and of Newton. However, after the publication of the latter's *Optiks* in 1704 little further progress was made during the eighteenth century in the fundamental understanding of the subject.

We can now see that the observations of Thomas Melvill (1726-53) were of great significance and that the experiments of this Scottish physicist, published in Edinburgh in 1752, make him one of the pioneers of spectrum analysis. He mixed various salts with burning spirits, and

'having placed a paste-board with a circular hole in it between my eye and the flame of the spirits, in order to diminish and circumscribe my object, I examined the constitution of these different lights with a prism (holding the refracting angle upwards) . . .

He observed the great predominance of the yellow light from salt which formed in the prism a sharp image of the hole through which the flame was viewed.

'Because the hole appears thro' the prism quite circular and uniform in colour; the bright yellow . . . must be of one determined degree of refrangibility;

and the transition from it to the fainter colour adjoining not gradual but immediate.'²

These important observations seem to have been largely ignored, although a few years later Marggraf appears to have been aware of the different colours imparted to flames by sodium and potassium salts and to have used them for analysis; and in 1822 Sir David Brewster introduced his monochromatic lamp using soda light, the idea of which, according to Roscoe³, was due to Melvill.

During the first two years of the nineteenth century both the infra-red and ultra-violet regions of the spectrum were discovered and in the course of verifying the existence of the latter, W. H. Wollaston, by the use of a very narrow slit for the first time, observed the discontinuous band spectrum of a candle flame visible through a prism. Newton had failed to see these lines because he had used light from a circular aperture and thus obtained a series of overlapping spectra. Wollaston wrote:

'When a very narrow line of the blue light at the lower part of the flame is examined alone . . . the spectrum, instead of appearing a series of lights of different hues contiguous, may be seen divided into five images, at a distance from each other. The 1st is broad green, terminated by a bright line of yellow; the 2nd and 3rd are both green; the 4th and 5th are blue, the last of which appears to correspond with the division of blue and violet in the spectrum.'⁴

Making further use of a fine slit, Wollaston was also able to discover these fixed *dark* lines in the solar spectrum.

Independently, these lines and very many others were observed by Josef Fraunhofer in 1814. This lens manufacturer and physicist made for himself a large prism, observed the double line (due to sodium) in flames from alcohol, sulphur, oil and candles, and made a careful study of the solar lines, mapping no less than 576 of them*.

Sir John Herschel investigated the spectra of numerous coloured flames, including those of salts of strontium, lithium and copper and of boracic acid, and remarked:

* Later, with a diffraction grating, Fraunhofer measured the wavelength of the D line. His mean value was 5888 Å. Ångström in 1868 found⁵ 5892.12 Å. The accepted modern value is 5896 Å (D₁) and 5889 Å (D₂).



Figure 1. (LEFT TO RIGHT) G. Kirchhoff, R. W. Bunsen and H. E. Roscoe in 1862—from *'Humor and Humanism in Chemistry'* by John Read (London: Bell, 1947)

'Of all the salts, the muriates succeed the best from their volatility . . . The colours thus communicated by the different bases to flame afford in many cases a ready and neat way of detecting extremely minute quantities of them.'

Similar observations were made by W. H. Fox Talbot, the pioneer in photography, and a few years later in 1834, Brewster studied the lines produced by sunlight passing through 'nitrous acid gas'. His stated aim was 'the discovery of a general principle of chemical analysis in which simple and compound bodies might be characterized by their action on definite parts of the spectrum'. He observed the effect of increasing the optical path length and the pressure of the brown oxide of nitrogen, which was in each case to make the lines in the red and yellow region more distinct and to broaden those in the blue and violet region. Similar

observations of absorption spectra were made about the same time by other workers including J. F. Daniell and W. H. Miller who passed light from a gas lamp through gaseous chlorine and bromine and observed it with a telescope after passing it through a prism. In the case of bromine more than one hundred lines were observed. The absorption spectra of a number of colourless gases were also studied by W. A. Miller in 1845.

Thus many workers before Bunsen and Kirchhoff had made observations of both emission and absorption spectra, and the analytical use of spectra had been appreciated. It remained however for the collaboration to show the full significance of the spectrum, to create a satisfactory experimental technique and to show its power in the discovery of new elements.

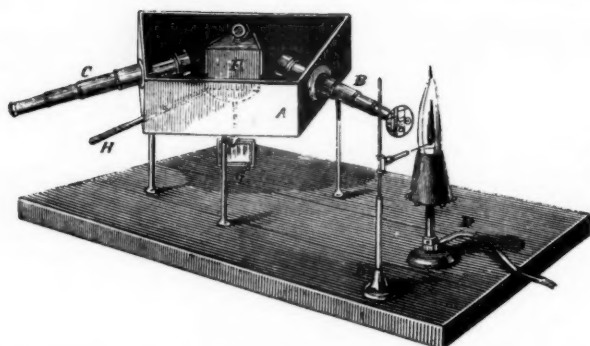


Figure 2. Bunsen's original form of spectroscope—from '*Spectrum Analysis*' by H. E. Roscoe (4th Edn, 1885)

R. W. Bunsen and G. W. Kirchhoff

Robert Wilhelm Bunsen (1811–99) was born at Göttingen and studied chemistry at that University under Strömeyer². At the age of twenty-five he succeeded Wohler as professor of chemistry at the higher technical school at Cassel and there carried out his earliest work in organic chemistry. This concerned the obnoxious and poisonous compounds derived from cacodyl oxide $(\text{CH}_3)_4\text{As}_2\text{O}$ and was of great importance at the time in supporting Berzelius' theory of radicals, but afterwards Bunsen took no part in either the practical or theoretical aspects of organic chemistry and would not even permit such work to be carried out in his laboratory. Subsequently he was professor at Marburg and Breslau, and finally at Heidelberg; all his later work was in inorganic chemistry and in physical chemistry and of the latter he must be regarded as a founder.

The study of light was a topic of great interest to Bunsen. Having observed that burning magnesium gave a light of very considerable intensity he invented his grease spot photometer with which to measure it, and finding the alcohol wick flame and the new English gas burners equally unsatisfactory, proceeded to design his own famous burner (1855). Two years before this he had begun investigations of photochemistry with the young Henry Roscoe. They proved that the amount of photochemical change is directly proportional to the amount of light absorbed and also investigated the induction period in the photochemical reaction of hydrogen and chlorine. This collaboration continued for some time, even after Roscoe's return to England, but it suddenly ended in 1859, due to the new work with Kirchhoff. Bunsen wrote to Roscoe on November 15th, 1859:

'At present Kirchhoff and I are engaged in a common work which doesn't let us sleep.'

Gustav Robert Kirchhoff (1824–87) had become acquainted with Bunsen at Breslau, and was frequently spoken of as Bunsen's greatest discovery of that period, for it was at the latter's recommendation

that Kirchhoff was appointed professor of physics at Heidelberg. Trained as a physicist, he was of course well aware of the earlier work on the spectrum. Now, it seems that Bunsen had been for some time making analytical use of the colours imparted by various salts to the colourless flame of his burner, and to distinguish them he viewed them through coloured glasses or solutions. Discussing this procedure with Kirchhoff, the latter pointed out that it would be better to pass the light from these flames through a prism and to observe the spectrum. Such was the origin of their collaboration. To it Bunsen brought a wide knowledge of chemistry and its practical techniques together with great manipulative ability; Kirchhoff on the other hand was more speculative and with greater mathematical ability was able to provide the necessary theoretical basis.

The earliest form of spectroscope employed by Bunsen is shown in Figure 2. In this the hollow prism *F* was placed in a blackened box, to one side of which was fixed a telescope *C*, whilst on another side was a tube having a slit at one end and a lens at the other. This form of *collimator* to render the rays parallel and obtain a pure spectrum had been proposed by Professor Swan³ of St. Andrews in 1855.

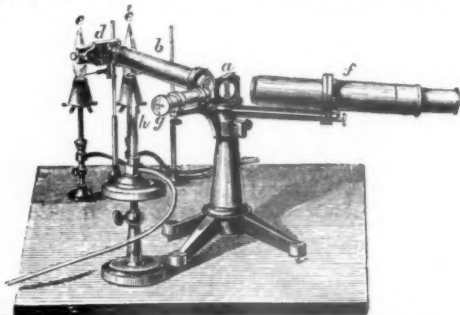


Figure 3. Revised form of spectroscope described by Bunsen and Kirchhoff—from '*Spectrum Analysis*' by H. E. Roscoe (4th Edn, 1885)

in fact all the essentials of the apparatus had been previously used or suggested; here, however, they were brought together for the first time. *Figure 3* shows the more refined instrument⁶ described by Bunsen and Kirchhoff in 1861 and made for them by C. A. Steinheil of Munich. In this two flames were used to facilitate comparison of spectra for analytical purposes; another flame was used to illuminate a mm scale in the tube *g* so that it was visible through the telescope between the two superimposed spectra.

so that the spectrum contained a dark line at that place⁴.

It was thus clear that hot gases absorbed the same kind of light as they emitted and Kirchhoff gave a mathematical proof⁷ that the ratio between the emissive and absorptive powers is the same for all substances at the same temperature for light of the same wave length. The Wollaston and Fraunhofer dark lines were thus a consequence of absorption by incandescent vapours and this put solar and stellar

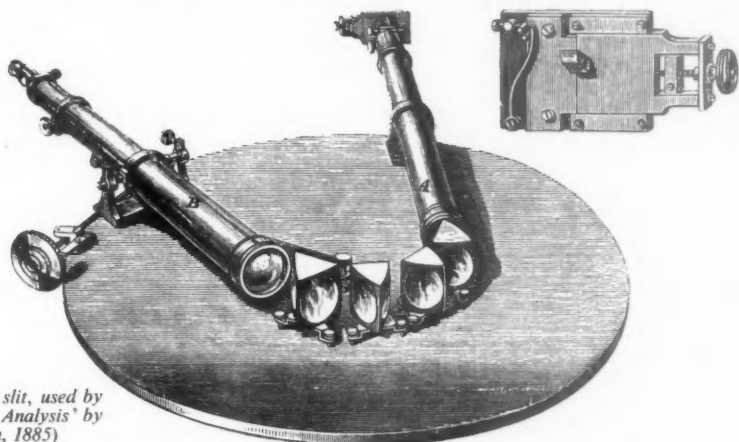


Figure 4. Spectroscope and slit, used by Kirchhoff—from 'Spectrum Analysis' by H. E. Roscoe (4th Edn, 1885)

A method of mapping the spectra using this scale was worked out by Bunsen⁸. A further improvement was made by Kirchhoff in his own instrument (*Figure 4*), namely the use of four prisms to obtain greater dispersion. A small reflecting prism was placed on the slit of the instrument to obtain the two superimposed spectra.

The first result of the new work was Kirchhoff's explanation of the relation between the bright and dark lines observed by Wollaston and Fraunhofer. On introducing common salt into the flame Bunsen and Kirchhoff observed that a yellow line appeared in the spectrum in exactly the position of a dark line in the solar spectrum. Attempting to observe both these lines simultaneously by allowing sunlight and yellow sodium light to shine on the slit of the spectroscope together, they were surprised to find, not that the dark line became yellow, but that it became darker. Kirchhoff showed that the dark line could be produced if a luminous flame, giving a continuous spectrum, was used as a source and the sodium flame introduced between this source and the slit. In explanation he gave the analogy of sympathetic vibrations: the white light from the luminous source lost those vibrations corresponding to the yellow line when it passed through the sodium flame,

spectroscopy on a firm basis, whilst Bunsen and Kirchhoff's instrument became a powerful new tool for studying the chemical composition of solar and stellar atmospheres. Bunsen's letter to Roscoe mentioned above continued:

'Thus a means has been found to determine the composition of the sun and fixed stars with the same accuracy as we determine sulphuric acid, chlorine, etc., with our chemical reagents.'

In their first paper⁹, however, Bunsen and Kirchhoff were concerned only with terrestrial spectra. They showed the extreme delicacy of the spectrum reaction of sodium, previously remarked upon by Swan, and the wide distribution in nature of lithium was revealed. It was shown how the presence in a mixture of two or all of the chlorides of sodium, potassium, lithium, calcium, strontium, and barium could be detected by their spectra using less than a milligram of material. They clearly established that 'the positions which the lines occupy in the spectrum are due to a chemical property as invariable and of as fundamental a nature as the atomic weight, and can therefore be determined with almost astronomical accuracy'—a fact which no previous workers had stated with certainty. Finally, Bunsen and Kirchhoff remarked that

'the method of spectrum analysis may also play

a no less important part as a means of detecting new elementary substances . . . We have had opportunity of satisfying ourselves that in reality such unknown elements exist.¹

New Elements

Shortly after the first experiments with the spectroscope, Bunsen was examining the alkalis left from the evaporation of a large quantity of mineral from Dürkheim. After removal of lime, strontia, magnesia and lithia, the residue showed the characteristic lines of sodium, potassium and lithium but in addition two remarkable blue lines previously unobserved. Bunsen and Kirchhoff concluded that there was present a new element belonging to the group of alkali metals and Bunsen set out to evaporate some forty tons of the mineral water to obtain it. By November 6, 1860, he was writing to Roscoe:

'I have been very fortunate with my new metal. I have fifty grams of the almost pure chloroplatinate, which I can easily make absolutely pure . . . I shall name it *caesium* because of its beautiful blue spectral line.'²

Caesium, then, was the first element to be discovered by means of the spectroscope. Its properties were found to be very similar to those of potassium; it had in fact been obtained as the sulphate in 1846 by Plattner who had mistaken it for a mixture of potassium and sodium sulphates. A few months later, in February, 1861, Bunsen and Kirchhoff were able to announce the discovery of yet another alkali metal, *rubidium*, obtained from the mineral lepidolite from Saxony, and characterized by two violet lines in its spectrum.

Other workers were quick to make further use of the new method. Soon after Bunsen and Kirchhoff's announcement of the discovery of rubidium, William Crookes was examining for tellurium some residues from a sulphuric acid plant, after removal of selenium. The spectroscope revealed no lines indicative of tellurium but did reveal a new bright green line which Crookes showed was due to a new element which he named *thallium*. The element was isolated by him in 1862 and also by C. A. Lamy who showed its metallic nature. Again, a year later, F. Reich and H. T. Richter of the Freiberg School of Mines discovered the existence of *indium* in some zinc ores, by means of a brilliant indigo line in its spectrum. The spectroscope was thus rapidly becoming indispensable equipment for the chemist and several workers devoted themselves to investigations with it. These included P. É. Lecoq de Boisbaudran whose discovery (1874) with its aid of yet another new element, *gallium*, in zinc blende was of great importance since this element filled a gap ('eka-aluminium') which Mendeléeff had left in his periodic table. The remarkable agreement between

the observed properties and those which Mendeléeff had predicted provided strong experimental support for his classification.

Within a few years five new elements had thus been discovered with the aid of the spectroscope and that number was to be steadily increased with its use in the investigations of the rare earths³. Bunsen himself made studies of the spectra of erbium, dysprosium and their compounds. He also used (1875) the spark discharge from an induction coil to obtain emission spectra when a higher temperature was required (as with gallium for example). In later work Kirchhoff mapped the lines of thirty-two elements and made a study of sun spots. Amongst the earliest spectroscopic studies of the sun were the independent observations by N. Lockyer and P. J. C. Janssen in 1868 of the prominences in the chromosphere in which a yellow line D_3 was observed very close to the D lines of sodium. Finding that this line did not belong to any known element Lockyer attributed it to an element *helium*; twenty-seven years elapsed before Ramsay discovered this element on the earth.

After the publication of the first papers by Kirchhoff⁴ and by Kirchhoff and Bunsen⁵, claims were made that others had anticipated their work. Kirchhoff showed¹⁰ that though the idea of 'chemical analysis by spectrum observations' had been proposed, it had not been satisfactorily realized in practice. Previous work¹ had suffered from two faults: bright line spectra free of the continuous spectra had not been used, and all of the sources had been contaminated with sodium. Kirchhoff's discovery of the relation between bright and dark lines clarified all this, and the collaboration of Bunsen and Kirchhoff placed the analytical use of spectra on a firm theoretical and experimental basis and demonstrated this in a striking way in the discovery of new elements.

References

- PEARSON, T. H. and IHDE, A. J. *J. chem. Educ.* (1951) **28**, 267
- MELVILLE, T. *Edinburgh Essays and Observations, Physical and Literary*, Vol. II. Quoted in WOLF, A. *A History of Science, Technology and Philosophy in the 18th Century* p. 171 London: Allen & Unwin, 1952
- ROSCOE, SIR HENRY E. *Spectrum Analysis* (4th Edn) London: Macmillan, 1885
- WEEKS, M. E. *Discovery of the Elements*, Chapter 20 Easton, Pa.: Journal of Chemical Education, 1948
- ROSCOE, SIR HENRY E. *Chemical Society Memorial Lectures, (1893-1900)* p. 530 London: Gurney & Jackson, 1901
- KIRCHHOFF, G. and BUNSEN, R. *Phil. Mag.* (1861) **22**, 329, 498
- *ibid.* (1860) **20**, 1
- and BUNSEN, R. *ibid.* p. 89
- GREEN, J. H. S. *Research, Lond.* (1958) **11**, 376
- KIRCHHOFF, G. *Phil. Mag.* (1863) **25**, 250

ORGANIC SEMICONDUCTORS

D. D. ELEY

Chemistry Department, University of Nottingham

This paper reviews results on the electrical conductivity of crystalline organic substances. Many substances behave as semiconductors with an energy gap which decreases with the number of π -electrons in the molecule. It is suggested that conductivity is associated with the intermolecular tunnelling of thermally excited π -electrons. The subject has implications in biology. This is one of the articles in the current semiconductor series; another article appears on page 300 and the last article will be published in the next issue of *Research*.

BY ORGANIC semiconductors we mean pure organic compounds or molecular complexes, more especially crystalline solids, showing a specific conductance κ which increases with temperature according to the well known relation

$$\kappa = \kappa_0 e^{-\Delta\epsilon/2kT} \dots (1)$$

We shall include in our survey substances with very small conductances at room temperature, e.g., with specific conductances of 10^{-15} ohm $^{-1}$ cm $^{-1}$ and energy gaps $\Delta\epsilon$ as high as 3 eV, although such substances would be regarded as insulators in industrial practice. However, the practical dividing line between semiconductors and insulators is of less interest at present, than the actual values of κ_0 and $\Delta\epsilon$ found for a wide range of organic compounds, and their relation to chemical structure. Values of $\Delta\epsilon$ are now well established in the range 0.1 to 3 eV and a representative choice of data are shown in Table 1. With one exception these figures are taken from work carried out in the author's laboratory.

The conductivity at a given temperature decreases very strongly as one descends Table 1. A comparison may be made with energy gaps for the elements¹, for example, grey tin 0.1 eV, germanium 0.74 eV and sulphur 2.6 eV, the latter being considered a typical insulator at room temperature. The earlier studies on electrical conductance in solid organic substances were concerned with rather complex photoconduction phenomena²⁻⁷.

Studies of organic substances as intrinsic semiconductors started in 1948 with independent publications by D. D. ELEY⁸ and A. T. VARTANYAN⁹ on phthalocyanine (Figure 1, LEFT) and its metal derivatives. Since phthalocyanine is well known to be one of the most heat-stable of organic molecules, it is not surprising that both authors chose the same substance for their first investigations. While Eley obtained an energy gap of 2.39 eV, Vartanyan's value was 0.87 eV, these values being respectively above and below what we now know as the true value. H. AKAMATU and H. INOKUCHI¹⁰ measured the semiconduction of isodibenzanthrone (Figure 1,

RIGHT) and two similar compounds, finding $\Delta\epsilon$ values of 0.75 eV.

The original suggestion was that the semiconductivity was associated with motion of π -electrons in a conduction band⁸, and this view has been borne out by most of the subsequent investigations of semiconductance, although the band is probably very narrow, about 0.1 eV).

Intrinsic Nature of the Conductivity

Discrepancies in the initial data may be attributed to failure to realize the effects of intercrystalline contacts, polymorphic transitions and impurities. We may illustrate this point by reference to metal-free phthalocyanine. Eley's original $\Delta\epsilon$ of 2.39 eV was obtained using d.c. with crystals lightly packed between platinum electrodes under hydrogen gas, over the range 300° to 470°C. The energy gap is

Table 1. A range of organic semiconductors—all substances measured in the dry state in vacuo

Substance	$\Delta\epsilon$ (eV)	$\log_{10} \kappa_0$ (ohm $^{-1}$ cm $^{-1}$)
Diphenyl picryl hydrazyl	0.26	6.1
Dimethyl aniline-bromanil complex	0.45	6.7
Isodibenzanthrone	0.96	2.3
Phthalocyanine	1.49	0.9
Haem	1.74	2.8
Thymus nucleic acid	2.44	3.45
Anthracene	2.7 *	2.0
Haemoglobin	2.75	4.6
Polyglycine	2.99	6.3

* see reference 19

here much increased over the true value by the effects of intercrystalline resistances. Measurement of the powder in a steel cell¹¹ under a pressure of 80 kg/cm² gave a value of 1.1 eV; in addition it was found that a change of slope occurred at 270°C, and below this the value of $\Delta\epsilon$ was 0.8 eV.

Figure 2 shows the resistance of the cell R related to κ by κ equals β/R , where the cell constant β

depends on the cell dimensions and a filling factor for the powder¹¹. A subsequent study was made using an a.c. method to eliminate contact resistance, referred to later¹². Four runs were made on three samples, each previously heated for a period at 350°C to complete the polymorphic transition. The four $\Delta\epsilon$ values all fell in the range 1.49 ± 0.03 eV (150° to 350°C). A change of slope was still found

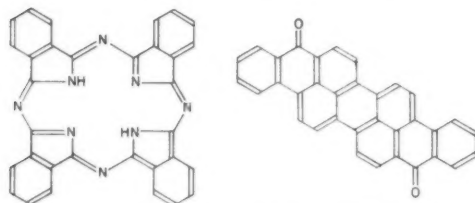


Figure 1. (LEFT) Phthalocyanine and (RIGHT) isodibenzanthrone

but at the lower temperature of 150°C, the exact point being associated with the trace metal content of the particular preparation as determined by spectroscopic analysis. The value of 1.49 eV may be compared with further values of $\Delta\epsilon$, in electron volts, for metal-free phthalocyanine:

1.71	d.c. with single crystals ¹³ ,
1.61–1.65	photoconductivity ¹⁴ ,
1.44–1.87	d.c. with single crystals ¹⁵ .

Therefore a wide range of methods and specimens give values in the range 1.5 to 1.7 eV. Vartanyan's original very low value of 0.87 eV was obtained over the relatively low temperature range of 10° to 100°C, and probably refers to an impurity semiconductivity. The impurity centres may in fact be dislocations as in this work the specimen was a thin film rubbed into quartz between gold electrodes.

Another substance which has been investigated in considerable detail is isodibenzanthrone. This substance (Figure 1) is much more easily purified by sublimation than phthalocyanine and in fact no trace metals were shown on spectroscopic analysis. The original $\Delta\epsilon$ of Akamatu and Inokuchi¹⁰ was 0.75 eV, obtained by using d.c. on the crystalline powder under a pressure of 80 kg/cm². A repetition of this result in our laboratory¹¹ using a steel compression cell gave a value of 1.14 eV; a determination using the a.c. method¹² gave 0.96 eV, which agrees well with the photoconductivity energy gap determined by Inokuchi¹⁰, $\Delta\epsilon = 0.93$ eV, and may be taken as the correct value.

There is still a need for more work on anthracene, as energy gaps of 1.65 eV, 1.92 eV, and 2.7 eV have been reported^{11,17–19}, the last being most

probably the correct value¹⁹. Other more complex molecules which have been studied by different workers using different techniques in the author's laboratory are diphenyl picryl hydrazyl and, at the other extreme of the scale, haemoglobin. In all cases the agreement in conductivity and energy gap is good enough to convince us that the effect must be an intrinsic property of the pure substances. Apart from metal-free phthalocyanine, the only substances to date which have shown a kink in the graph of $\log \kappa$ (or $\log R$) against temperature, are certain amino acids, where the discontinuity occurs around 127°C. We have not yet investigated the relative roles of impurities and polymorphic transitions for the amino acids.

Measuring Techniques

The techniques used in the author's laboratory at present are mainly d.c. measurements on crystalline powders between platinum electrodes under a pressure of about 80 kg/cm² *in vacuo* or a.c. measurements on crystalline powders between platinum electrodes under light compression *in vacuo*. The

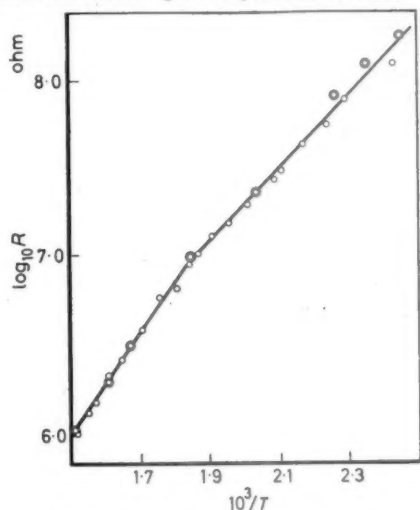


Figure 2. The resistance-temperature characteristic for metal free phthalocyanine under 80/kg cm²—●—ascending temperature; ○—descending temperature

first technique is used for all the higher resistance specimens, such as proteins, and measurements have been carried out up to 10^{15} ohm using a valve voltmeter with an ET1 electrometer triode. Figure 3 shows the results of a series of experiments on haemoglobin carried out by Dr M. H. CARDEW using this method²⁰. Generally speaking, organic crystals

obey Ohm's law up to a potential gradient of about 1000 volt per centimetre. The measuring cells contain certain electrodes mounted on silica supports^{12,20} to lower the 'blank conductance'.

The a.c. method depends on the fact that the equivalent circuit of a powder may be regarded as a resistance ρ_3 (of the crystals themselves) in series with a parallel arrangement of condenser C_2 and resistance ρ_2 (typical of the intercrystalline gaps). A measurement of empty and full conductivity cells is taken at a range of frequencies between 30 kc/sec and 30 Mc/sec, thus giving a value for the powder. Figure 4 shows a plot due to DR G. D. PARFITT of the resistance of isodibenzanthrone which is seen to fall to a constant value ρ_3 at high frequencies characteristic of the crystal. In Figure 4, ρ_3 values are plotted according to Equation 1 and yield $\Delta\epsilon$ equal to 0.96 eV.

Measurements of d.c. conductivity have also been made on single crystals^{13,15,17,19} and this technique allows measurements to be made along different crystal axes. In anthracene the specific conductance was equal along the *a* and *b* axes and nearly twice that along the *c* axis¹⁷.

The 'photoconductivity energy gap' may be obtained by two standard methods¹:

- (i) the value of the quantum energy where the spectral sensitivity has fallen to half value, $\epsilon_{\lambda\frac{1}{2}}$ on the long wave side; and

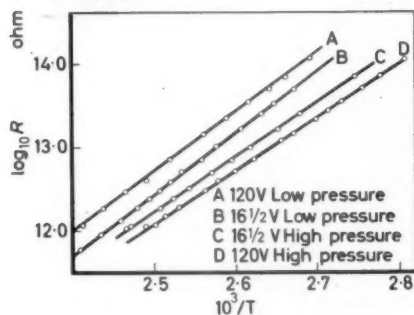


Figure 3. Resistance-temperature graphs for haemoglobin

- (ii) from the temperature variation of the photoelectric current, by an equation analogous to Equation 1, ϵ_ϕ .

These two methods give similar results for phthalocyanine¹⁴. A third method has been used¹⁶ to measure the long wave threshold quantum value, $h\nu$ of the spectral response curve (which is of course always lower than $\epsilon_{\lambda\frac{1}{2}}$) and finds it closely approximates to the thermal energy gap $\Delta\epsilon$ for isodibenzan-

throne and nine similar polyaromatic compounds. Further investigations are desirable, as photoconductivity in organic substances is a more complex phenomenon than semiconductivity, the results

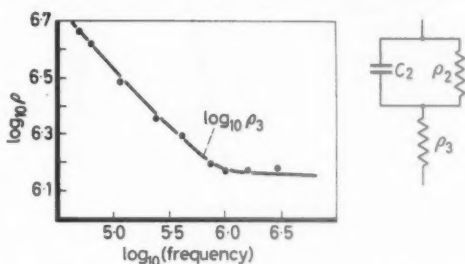


Figure 4. (LEFT) Resistance-frequency plot for isodibenzanthrone; (RIGHT) equivalent circuit

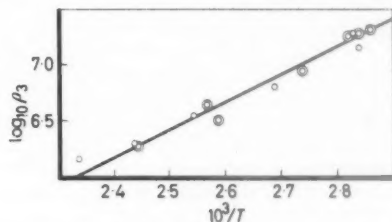


Figure 5. The high frequency resistance (ρ_3) for isodibenzanthrone as a function of ascending \odot and descending \circ temperature

observed depending very much on the geometry of the specimen and its attached electrodes. Thus D. J. CARSWELL²¹ and L. E. LYONS²² have found the photocurrent in anthracene crystals reproduced the absorption spectrum of the crystals; J. FERGUSON and W. G. SCHNEIDER²³ have emphasized that this result is found only when the crystal is illuminated on the opposite side to the face bearing the spot electrodes.

Few measurements have been made of thermoelectric power, and none of the Hall effect in organic substances owing to the high specific resistivities involved.

The π -Electron Hypothesis

It was immediately noticed that the effect of replacing the two central hydrogen atoms in phthalocyanine by metals such as copper, magnesium, zinc, etc had a very small effect on the conductivity and the energy gap^{8,9}, so that the conductance might be regarded as an integral part of the conjugated ring structure, most probably involving the π -electrons. Inokuchi²⁴ found that conductivity increased and $\Delta\epsilon$ decreased over a range of eleven polynuclear hydrocarbons

with the number of π -electrons in the molecule, and a similar result holds for polyazoaromatic ring systems²⁵.

The following model has been suggested^{11,12}, which accounts broadly for the results observed.

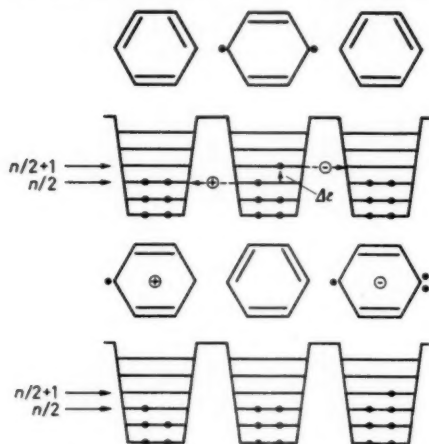


Figure 6. A simple model for organic semiconductors

Reference to Figure 6 shows three aromatic molecules, benzene for simplicity, side by side in the crystal. This diagram is only illustrative and not meant to convey the actual packing in the lattice. The work of J. MONTEATH ROBERTSON on anthracene, phthalocyanine and similar compounds has shown that the planar molecules lie at an angle to the (*ac*) plane, alternate molecules pointing above and below the plane. The molecules are represented by potential energy wells, the six π -electrons filling up the lowest three of six molecular orbitals common to the whole molecule, as in the left hand ring.

To calculate the energy of these orbitals we use the simplest of all models²⁶⁻²⁸, the perimeter model. The electrons form waves in a molecular path of length nd , where n is the number of π -electrons and d equals 1.39 Å is the C-C distance in aromatic compounds. Two sets of levels are considered: when the electron path forms a closed ring ('closed path') and, secondly, where the electron path is not closed ('open path'). The case of an unexcited molecule is shown on the left (Figure 6). The conduction process is separated into two steps.

(i) the electron is excited from the highest filled $n/2$ th level to the lowest unfilled $(n/2 + 1)$ th level. The energy gap (full arrow, Figure 6) is

$$\Delta\epsilon = \frac{h^2}{8md^2} \frac{(n+1)}{n^2} \quad (\text{open path}) \dots (2)$$

$$\Delta\epsilon = \frac{h^2}{4md^2} \cdot \frac{1}{n} \quad (\text{closed path}) \dots (3)$$

(ii) the excited electron may then tunnel at the $(n/2 + 1)$ th level to the next molecule with a negligible energy requirement (dotted arrow, Figure 6). It may also be possible for an electron to tunnel at the $n/2$ th level from the left hand molecule to fill the empty space in the centre molecule, and this is equivalent to a 'positive hole' passing in the opposite direction as shown. We expect the mobility of the charge carrier to be determined mainly by the probability of tunnelling through the intermolecular barrier. Since the barrier will become thinner towards the top we should expect the electron (at the $(n/2 + 1)$ th level) to have a greater mobility than the positive hole (at the $n/2$ th level).

Figure 7 shows that broadly, the theory gives, an account of the observed facts. Here we plot experimental data for a wide range of π -electron compounds and most of the experimental points fall in between the $\Delta\epsilon$ - n relations computed for the open and closed paths, Equations 2 and 3. The figure contains all the known data to 1955; there are now some further points which may be added²⁹. The number of mobile π -electrons n , is of course not necessarily the same as the total number in the molecule, a point already made³⁰ for phthalocyanine on the basis of the observed diamagnetism. Some points may lie off the graph for this reason.

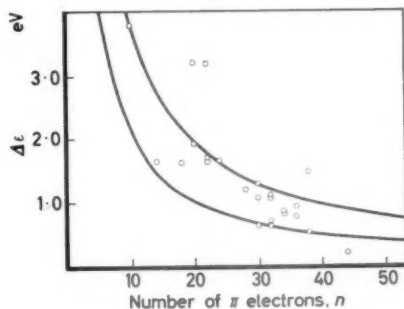


Figure 7. Thermal energy graph $\Delta\epsilon$ as a function of number of π -electrons, n . Comparison of experimental points with theory; upper curve, closed and lower curve, open path

On molecular orbital theory a free radical with $(n + 1)$ π -electrons may be regarded as possessing a single unpaired electron in the $(n/2 + 1)$ th level. Therefore if the intermolecular tunnelling requires no activation energy we should expect the energy gap for a solid free radical to be zero. Measurements established a very small energy gap, $\Delta\epsilon$ equals 0.26 eV

for diphenyl picryl hydrazyl crystals (see Figure 8), which gives strong support for this theory¹²; a recent repetition³¹ of this experiment using an evaporated film of diphenyl picryl hydrazyl between evaporated aluminium electrodes gave a closely similar value, 0.16 eV. The actual energy barrier between molecules will depend very much on molecular packing, as is shown by differences in conductance between the alpha and beta forms of phthalocyanine¹². It must be concluded that the molecular stacking in diphenyl picryl hydrazyl is particularly favourable for electron transfer.

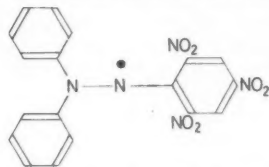


Figure 8. Diphenyl picryl hydrazyl

We have not so far distinguished the electron spins in the $n/2$ and $(n/2 + 1)$ states of the excited molecule. If antiparallel the excited state is a singlet state. If the electron spins are parallel the excited state is a triplet state, which is always lower in energy than the corresponding singlet state (with the same orbital wave function). In a first comparison¹¹, it was noted that the $\Delta\epsilon$ values for anthracene and phthalocyanine were considerably less than the energy values for the absorption maxima—3.33 eV and 1.75 eV respectively—of these molecules in solution.

With current values for the energy gaps of 2.7 eV and 1.49 eV the discrepancy is less marked and probably would not exist at all, if threshold optical energies were taken (see below). The suggestion had been made that the energy level for the $(n/2 + 1)$ molecular orbital was considerably broadened in the solid, by molecular interaction in the lattice. Even with the new figures one would require a band width of 1.32 eV in anthracene and 0.52 eV in phthalocyanine to explain the discrepancy, and such a width seems very large. D. C. NORTHROP and O. SIMPSON¹⁸ have obtained $\Delta\epsilon$ values for anthracene (1.93 eV) and pyrene (2.02 eV) which correlate closely with the calculated values of the ground singlet \rightarrow first triplet excitation energy. However, higher $\Delta\epsilon$ values were obtained by Inokuchi, namely 2.7 eV for anthracene and 2.4 eV for pyrene. Also, identification of the triplet and semiconducting states would conflict with the close correlation of $\Delta\epsilon$ with photoconduction threshold and long wave absorption edge noted for a number of polyaromatic compounds¹⁸ since the singlet \rightarrow

triplet transition is forbidden and not normally observed spectroscopically.

There remains a contradiction in the present comparison of thermal and optical excitation energies which can only be resolved by further data on $\Delta\epsilon$ values, such as that for anthracene, and a consideration of the theoretical significance of long wave optical thresholds in organic molecules.

Molecular Complexes

Semiconducting properties for complexes of alkali metals with anthracene with composition variable in the range M_1An to M_2An have been described by W. A. HOLMES-WALKER and A. R. UBBELOHDE³². Energy gaps, as defined in this paper, are $Li_{1.16}An$ 2.68 eV, $Na_{1.08}An$ 2.40 eV, $K_{1.18}An$ 2.20 eV. Increasing the concentration of metal atoms lowers the $\Delta\epsilon$ value. From one viewpoint these systems may be regarded as *n*-type impurity semiconductors with very high impurity (metal) concentrations. Clearly, the donor level of the metal atoms in the more dilute systems must be close to the top of the 'valency band' of the anthracene, and it would be interesting to see if a change to intrinsic anthracene conductivity ($\Delta\epsilon = 2.7$ eV) could be observed in favourable circumstances. The structure of these complexes has been further discussed³³.

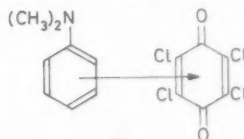


Figure 9. Dimethyl aniline chloranil

Akamatsu and co-workers reported on the properties of dark coloured chemically unstable complexes formed between polycyclic aromatic hydrocarbons and halogens³⁴. These showed a relatively high conductivity with $\Delta\epsilon$ values of 0.1 to 0.2 eV. Inokuchi and Eley³⁵ have examined complexes of the type dimethyl aniline chloranil (Figure 9). Other work had established that similar complexes of tetramethyl *p*-phenylene diamine possess a degree of biradical character, as shown by electron resonance signals³⁶. The complexes were found to possess an appreciable semiconductivity, which remained approximately constant as the acceptor was changed over the series chloranil, bromanil, iodanyl; the paramagnetism, actually surface paramagnetism measured by the parahydrogen conversion, increased strongly over the series.

Rose-Innes (in a private communication) had also observed semiconductivity in the tetramethyl *p*-phenylene diamine chloranil complex.

The paramagnetism may be taken to be proportional to the number of unpaired electrons in molecular orbitals, and the results show that while the crystals have a relatively high conductivity as we would expect, this conductivity is limited by some factor besides the number of unpaired spins. It is hoped to report more extensive data in the near future³⁷. The semiconductivity of molecular complexes may be important in biological problems, for example, chlorophyll-protein complexes in photosynthesis^{38,39} and carcinogenic hydrocarbon-protein complexes⁴⁰.

Proteins

Speculations on the role of electron migration in proteins have been made by various workers⁴¹⁻⁴³ and in one review⁴⁴ it was calculated that the repeat unit



will lead to three bands of width 0.2 eV, the lower two being filled with electrons and the upper band empty. The energy gap, for sp^2 hybridization of the N atoms, was calculated as 3.05 eV. Experiments on fibrous⁴⁵ and globular proteins¹¹ in the dry state gave a series of $\Delta\epsilon$ values in the range 2.2 to 2.7 eV. Recently we have completed a systematic study²⁰ finding, for example, values (all in the dry state) for globin 2.97 eV, haemoglobin 2.75 eV, glycine 2.92 eV, and polyglycine 3.12 eV; similar data have recently been found by D. SPIVEY in the author's laboratory for a range of dry proteins.

Since polarization effects are small we regard this conductivity as electronic, and probably originating in the hydrogen bridge system. It is possible that in wet proteins a proton conductivity occurs, a possibility discussed theoretically by K. WIRTZ⁴⁶ and favoured by N. RIEHL⁴⁷ for his results on wet gelatin. It is known that a light quantum ($h\nu$ equal to 4.43 eV) absorbed by tyrosine and tryptophane residues, can migrate to the iron atom and dissociate carbon monoxide from carbon monoxy-haemoglobin⁴⁸.

The above experiments suggest that 4.43 eV is more than adequate to excite an electron into the conduction band of the protein. The electron and positive hole so formed might then diffuse together, carrying the energy as an 'exciton'. Under the influence of an applied field the electron and hole will diffuse in opposite directions, a process which may give rise to processes of oxidation and reduction occurring at the opposite ends of a protein system. Interesting possible mechanisms have been discussed by T. A. GEISSMANN⁴⁹. The suggestion has been made that polyaromatic hydrocarbons may form a complex with a protein transferring an electron from the upper most filled protein band to an empty level of the hydrocarbon of the protein¹⁰. It was

assumed that for this process to take place an unfilled band in the aromatic molecule must occur at the same level as the band in the protein. This restricts the transfer process to a few hydrocarbons, which turn out to be the more carcinogenic molecules.

Semiconductivity and Catalysis

In catalysis by metallic oxides, it is known that chemisorbed gases transfer electrons to and from the catalyst. Thus an electronegative gas such as oxygen removes electrons from the solid, increasing the conductivity of p -type Cu_2O where the current carriers are positive holes⁵⁰, but decreasing that of n -type ZnO where they are electrons⁵¹. Electron donors (reducing agents), such as hydrogen and carbon monoxide have the reverse effects. Somewhat analogous results were found by A. PETRIKALN⁶ for organic substances, in that absorbed oxygen increases the photoconductivity after effect in films of eosin and erythrosine, but lowers it for triphenylmethane dyes. Further studies have been recently made by R. C. NELSON⁵², and the effects are certainly more complex than semiconduction in phthalocyanine and similar compounds.

Similar effects of electron transfer, but less marked, are observed when gases adsorb on metals⁵³, and we have recently sought for analogies with diphenyl picryl hydrazyl²¹. A film of this compound between aluminium film electrodes is unaffected in its conductivity if hydrogen, nitrogen or oxygen are admitted. If one of the electrodes is a film of palladium then hydrogen effects a decrease in electrical conductivity, and therefore presumably decrease in the number of conduction electrons in the diphenyl picryl hydrazyl. The conclusion was drawn that if hydrogen atoms are provided, as by the palladium film, these will form covalent links with the free radical electrons in the diphenyl picryl hydrazyl.

One function of a redox catalyst is to provide a pathway for the easy transfer of electrons from reducing to oxidizing agents. Certain proteins are known to play a specific role in redox catalysts, but unless the wet protein has a quite different electronic behaviour from the dry protein, it would seem the high energy gap of 2.7 eV makes the protein conductivity band inaccessible for this purpose, though it can obviously play a role in photochemical and radiation effects.

In molecules such as haemoglobin, cytochrome and peroxidase we have porphyrins attached to specific proteins and it is interesting to estimate how far the protein can chemically extend the mesomeric system of the protein. Reference to Figure 7 shows that for an open path an energy gap of 2.7 eV for the protein is equivalent to seven or eight π -electrons in

an aromatic system. This is the extension to be expected to the porphyrin system if the porphyrin ring is closely coupled to the protein. In fact, the red shift in the spectrum of prosthetic groups attached to proteins suggests that the equivalent extension of the conjugation system by the protein is even smaller, perhaps only two π -electrons²⁰. This effect would seem to be too small to bring about the profound alteration in reactivity observed in practice when a specific protein is coupled to a porphyrin or other group to form an enzyme.

The Mobility

The factor κ_0 for an intrinsic semiconductor is given by 4μ at 20°C, where μ is the average mobility of the charge carriers, electrons and holes¹. The mobility on the present model will be determined by the ease with which the charge carriers can pass from one molecule to the next in the lattice by the tunnelling process¹¹. It has been suggested⁵⁴ that this process will give rise to a relation between μ and so κ_0 and $\Delta\epsilon$, such as may be seen roughly to hold for the data of Table 1. It is probably that several relations of this type exist, depending on the mode of molecular packing, simple aromatic systems lying on one line and protein type systems on another²⁰.

Conclusions

The intrinsic nature of the semiconductivity is now established for many organic compounds, and energy gaps have been determined over a range 0.1 to 3.0 eV. Considerable problems remain in determining the relative mobility of electrons and positive holes and in relating photoconductivity to semiconductivity. Molecular complexes, and solid free radicals, furnish the most efficient organic semiconductors.

Apart from possible applications in the technical field, the study of organic semiconductors should help to throw light on enzyme catalysis, and on problems of energy transport in biological systems, where the formation of molecular complexes may play an important role.

In closing this review the author would like to acknowledge the enthusiastic collaboration of his research students, past and present, and particularly Dr G. D. Parfitt and Dr M. H. Cardew.

References

- Moss, T. S. *Photoconductivity in the Elements* London: Butterworths Scientific Publications, 1952
- Pochettino, A. *Atti. Acad. Lincei* (1906) **15**, 171, 355
- Volmer, M. *Z. Elektrochem.* (1915) **21**, 113
- Rabinowitsch, M. *Z. phys. Chem.* (1926) **119**, 59
- Petrikal, A. *ibid.* (1930) **10B**, 9
- Vartanyan, A. T. *Acta Physicochim. U.S.S.R.* (1947) **22**, 201
- Putzeiko, E. A. *Doklady Akad. Nauk U.S.S.R.* (1948) **59**, 471
- Eley, D. D. *Nature, Lond.* (1948) **162**, 819
- Vartanyan, A. T. *J. phys. Chem. U.S.S.R.* (1948) **22**, 769
- Akamatsu, H. and Inokuchi, H. *J. chem. Phys.* (1950) **18**, 810
- Eley, D. D., Parfitt, G. D., Perry, M. J. and Taysum, D. H. *Trans. Farad. Soc.* (1953) **49**, 79
- Eley, D. D. and Parfitt, G. D. *ibid.* (1955) **51**, 1529
- Fielding, P. E. and Gutmann, F. *J. chem. Phys.* (1957) **26**, 411
- Vartanyan, A. T. and Karpovich, I. A. *Doklady Akad. Nauk U.S.S.R.* (1956) **111**, 561
- Kleitmann, D. *Electrical Properties of Phthalocyanine* U.S. Dept. of Commerce, Office of Technical Services Report PB 111419 (1953)
- Inokuchi, H. *Bull. chem. Soc. Japan* (1954) **27**, 22
- Mette, H. and Pick, H. *Z. Physik* (1953) **134**, 566
- Northrop, D. C. and Simpson, O. *Proc. Roy. Soc.* (1956) **234**, 124
- Inokuchi, H. *Bull. chem. Soc. Japan* (1956) **29**, 131
- Cardew, M. H. and Eley, D. D. *Energy Transfer with Special Reference to Biological System* Faraday Society Discussion (Nottingham, April 1959)
- Carswell, D. J. *J. chem. Soc.* (1957) 3648, 3661
- Lyons, L. E. and Morris, G. C. *ibid.* pp. 3648, 3661
- Ferguson, J. and Schneider, W. G. *Canad. J. Chem.* (1958) **36**, 1633
- Inokuchi, H. *Bull. chem. Soc. Japan* (1951) **24**, 222
- *ibid.* (1952) **25**, 28
- Bayliss, N. S. *J. chem. Phys.* (1948) **16**, 287
- Simpson, W. T. *ibid.* (1948) **16**, 1124; (1949) **17**, 1218
- Kuhn, H. *ibid.* (1949) **17**, 1198
- Meier, H. *Z. phys. Chem.* (1958) **208**, 325
- Lonsdale, K. *Proc. Roy. Soc. A* (1937) **159**, 149
- Eley, D. D. and Inokuchi, H. *Z. Elektrochem.* (1959) **63**, 29
- Holmes-Walker, W. A., and Ubbelohde, A. R. *J. chem. Soc.* (1954) 720
- Slough, W. and Ubbelohde, A. R. *ibid.* (1957) 982
- Akamatsu, H., Inokuchi, H. and Matsunaga, Y. *Bull. chem. Soc. Japan* (1956) **29**, 213
- Eley, D. D. and Inokuchi, H. *Proc. Third Biennial Carbon Conference, Buffalo* (1957) London: Pergamon Press, 1959
- Kainer, H., Bül, D. and Rose-Innes, A. C. *Naturwiss.* (1954) **41**, 303
- Eley, D. D., Inokuchi, H. and Willis, M. R. Paper to be submitted to the Faraday Discussion (Kingston, Ontario, September 1959)
- Arnold, W. and Sherwood, H. K. *Proc. Nat. Acad. Sci.* (1957) **43**, 105
- Calvin, M., Tollin, G. and Sogo, P. B. *J. Chim. Phys.* (1958) **55**, 869
- Mason, R. *Nature, Lond.* (1958) **181**, 820
- Möglich, F. and Schön, M. *Naturwiss.* (1938) **26**, 693
- Jordan, P. *ibid.* p. 693
- Szent-Györgi, A. *Nature, Lond.* (1941) **148**, 157
- Evans, M. G. and Gergely, J. *Biochim. Biophys. Acta* (1949) **3**, 188
- Baxter, S. *Trans. Farad. Soc.* (1943) **34**, 207
- Wirtz, K. *Z. Naturforsch.* (1948) **36**, 131
- Riehl, N. *Koll. Z.* (1957) **151**, 66
- Bucher, T. and Kaspers, J. *Biochim. Biophys. Acta* (1947) **1**, 21
- Geissman, T. A. *Quart. Rev. Biol.* (1949) **24**, 309
- Garner, W. E., Gray, T. J. and Stone, F. S. *Proc. Roy. Soc. A* (1949) **197**, 314
- Morrison, S. R. *Advances in Catalysis* (1955) **7**, 259
- Nelson, R. C. *J. chem. Phys.* (1951) **19**, 798
- Suhrmann, R. *Advances in Catalysis* (1955) **7**, 303
- Many, A., Harnik, E. and Gerlich, D. *J. chem. Phys.* (1955) **23**, 1733

BISMUTH TELLURIDE AND RELATED COMPOUNDS

D. A. WRIGHT

Research Laboratories of The General Electric Company Limited, Wembley, England

This article summarizes the present knowledge of bismuth telluride and similar compounds, and of alloys formed by substitution of elements from the same column of the periodic table. Particular attention is paid to those properties which have a bearing on thermoelectric performance. This is one of the articles in the technology of semiconductors series; another article appears on page 293 and the last article on solid state devices will be published in the next issue.

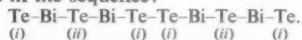
THE VERY active research on semiconductors in the last ten years has until recently been concentrated primarily on germanium and silicon. Work on semiconducting compounds was limited at the beginning of this period, and was largely aimed at the discovery of alternative materials for use in fabricating *p-n* junction devices. When the high mobilities of indium antimonide and indium arsenide were discovered, these materials rapidly became important for applications of the Hall effect and for infra-red detection because of their photoconductive properties. The present interest in a wide range of semiconducting compounds follows the realization that there may be important applications of semiconductors for thermoelectric refrigeration and generation. While in some respects the requirements for these applications are very exacting, in others they are much less so than for transistor applications. In particular there is much less need for ultra high purity and crystal perfection, so that materials of greater inherent complexity than germanium and silicon are potentially interesting.

In this laboratory, following the realization that semiconducting compounds of high mean atomic weight were of special interest for thermoelectric applications, it was decided to concentrate on bismuth telluride¹. The structure of bismuth telluride is more complicated than that of any other semiconductor which has so far received detailed study, and some of its physical properties are correspondingly different. In order to obtain an adequate understanding and control of this compound, our investigations have extended beyond those properties directly concerned with thermoelectricity. A fairly detailed study has now been made, and the intention of this article is to survey the present knowledge of bismuth telluride and some

other closely related compounds, referring specifically to the work done here in the Solid State Physics Laboratory. Short surveys of knowledge at an earlier stage have already been published²⁻⁵.

Structure and Physical Properties of Bismuth Telluride

Bismuth telluride, Bi_2Te_3 , forms a crystal structure of the rhombohedral class $R\bar{3}m$. The unit cell dimensions are: lattice parameter 10.47 Å and rhombohedral angle $24^\circ 9'$. Regarded as a hexagonal lattice the unit cell dimensions are⁶: *a* 4.384 Å and *c* 30.487 Å. Layers of atoms are stacked in the rhombohedral (111) direction in approximately cubic close packing, giving a repeating five-layer structure in the sequence:



Clearly the bonding between bismuth and tellurium atoms of type (i) is different from that between bismuth and tellurium of type (ii). It is probable that the bonding between the adjacent tellurium layers (i) is of the van der Waals type; it is certainly very weak, and is responsible for the ready cleavage along the basal (0001) planes, perpendicular to the axis of rotation, *i.e.*, the hexagonal *c* axis. The tellurium_(ii) atoms are surrounded almost octahedrally by six bismuth atoms, and each bismuth atom is surrounded almost octahedrally by six tellurium atoms, the spacing being 3.22 Å between bismuth and tellurium_(ii) atoms, and 3.12 Å between bismuth and tellurium_(i) atoms⁷. It is probable that the $\text{Bi-Te}_{(ii)}$ bonds are almost fully covalent, while there is some degree of ionicity in the $\text{Bi-Te}_{(i)}$ bonds.

The density of Bi_2Te_3 is 7.86, and the melting point 585°C . The coefficient of expansion is anisotropic, with a value at 20°C of 22.2×10^{-6} in the *c* axial direction and 12.9×10^{-6} in the *a* direction⁸. The specific heat at liquid helium

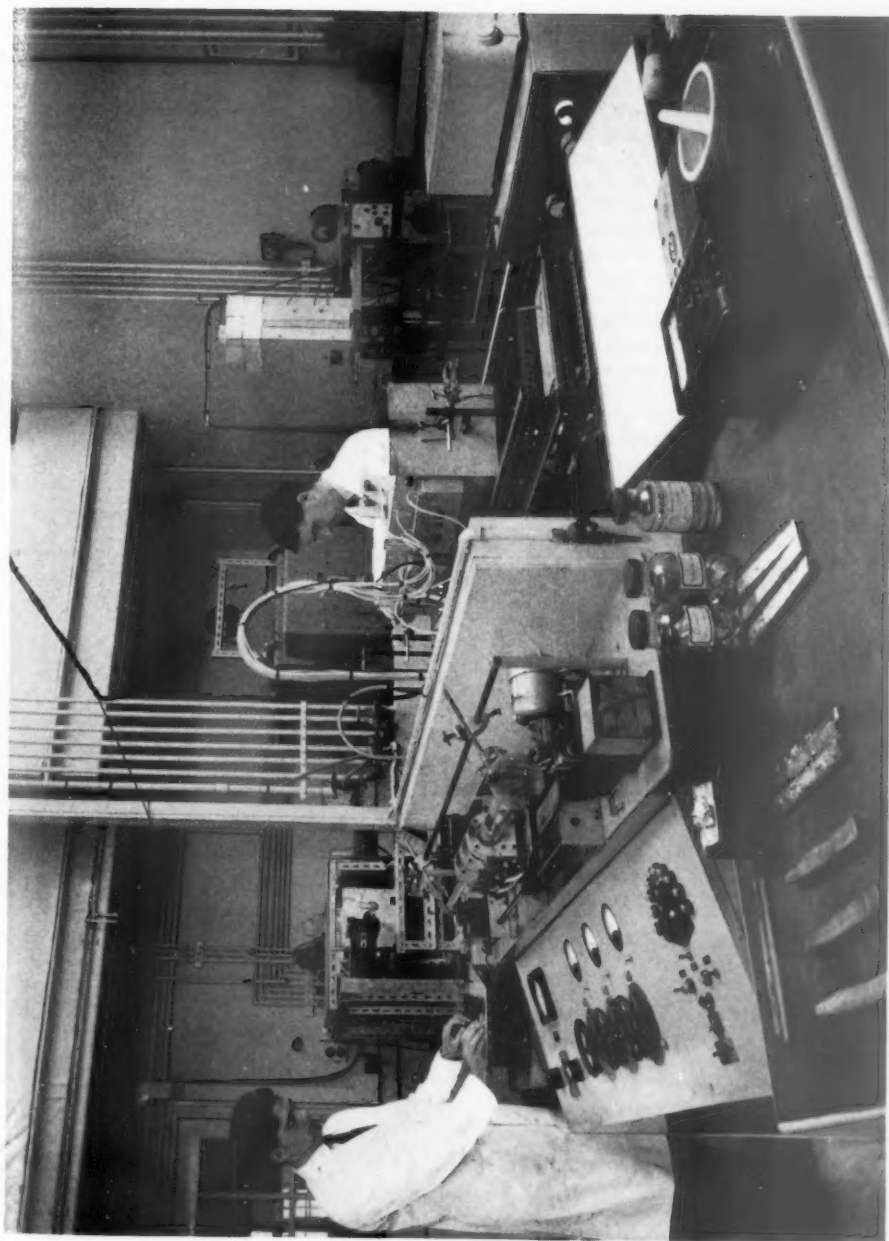


Figure 1. Apparatus used in zone-refining bismuth telluride

temperatures has been measured by C. A. BAILEY at the Clarendon Laboratory, Oxford; from this work a Debye temperature of 145°K is calculated.

Bismuth telluride and related materials are prepared from elements first melted together in the correct proportions, using materials with a purity of 99.99 per cent. The compound is then zone melted. *Figure 1* shows equipment used in this process by A. R. SHEARD⁸. With the simplest procedure, polycrystalline ingots are produced in which the crystals are orientated with the cleavage planes parallel to the direction of travel of the molten zone; the cleavage planes are not however parallel to one another. If a seed crystal is placed at the tapered end of a silica boat, an ingot which is initially polycrystalline can be zone melted from the region of contact with the seed, and a single crystal can be obtained. This is essential for the study of anisotropy. Typical dimensions obtained in this way are 12 to 16 cm \times 1.5 cm in the cleavage planes by 1 cm to 1.5 cm in the *c* direction.

Using stoichiometric proportions of bismuth and tellurium, and no deliberate additives, the compound

produced by zone refining is invariably *p*-type, with a carrier concentration near 10^{18} holes/cm³, and a resistivity of about 2×10^{-3} ohm cm for current flow along the direction of the cleavage planes. The resistivity is near 5×10^{-3} for current flow in the perpendicular direction. The material can be made more *p*-type with excess bismuth, or by addition of lead or cadmium, while it can be changed to *n*-type by using excess tellurium, or by adding a halogen, or some of the group I elements. As doping levels are increased it becomes more difficult to prepare single crystal material, so that the basic measurements on single crystals are so far confined to low levels of doping. With *n*-type material, the resistivity in the *c* direction is between four and five times that in the direction of the cleavage planes. Using about 0.06 per cent iodine, compensated material is obtained with the maximum resistivity value, which is near 7×10^{-3} ohm cm for current flow along the cleavage planes.

The single crystals produced by zone melting cleave readily to give a very flat surface. Etching reveals patterns as shown in *Figure 2*. There is

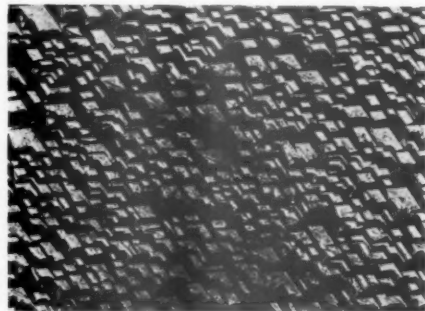
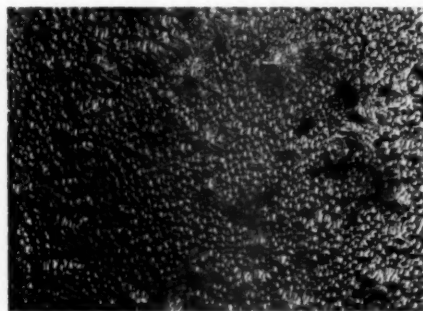


Figure 2. The cleavage-face of a (LEFT) p-type and (RIGHT) n-type Bi₂Te₃ crystal after etching

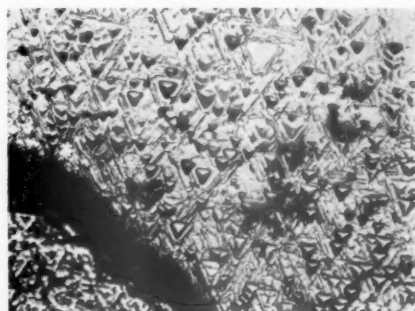
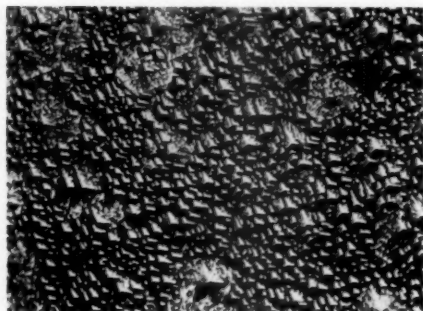


Figure 3. The etched surface of a p-type crystal after annealing at (LEFT) 450°C for 24 hours and (RIGHT) 500°C for 24 hours

some evidence that the etch pits are due to edge dislocations parallel to the c axis. Recent studies by J. Woods have shown that there is some reduction in etch pit density following annealing

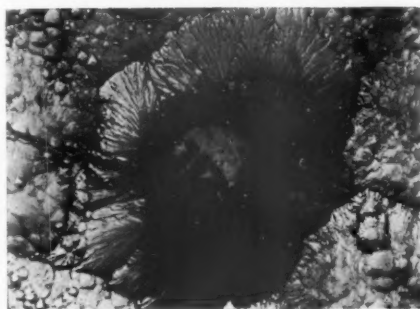


Figure 4. A circular p -type region with surrounding n -type in a nearly 'compensated' crystal with 0.06 per cent iodine

(Figure 3) and that in compensated material there are frequently p -type regions, segregated from n -type regions which etch differently. Furthermore, in the n -type regions there are frequently circular p -type areas; these sometimes surround triangular etch pits; Figure 4 shows a circular region in which the dark markings are due to probes used to identify the p - and n -type areas. These areas indicate the presence of cylindrical p - n junctions surrounding dislocation lines, and suggest that an impurity, e.g., iodine, has precipitated at the dislocation lines. These inhomogeneities would be undesirable for thermoelectric applications, tending to lower the electron mobility and to raise the apparent thermal conductivity. However compensated material is not used for such applications.

Requirements for Thermoelectric Applications

For a semiconductor to be useful for thermoelectric applications the figure of merit, which may be defined as

$$z = \alpha^2 \sigma / K \quad \dots (1)$$

should be as high as possible. Here α is the thermoelectric power, σ is the electrical conductivity and K the thermal conductivity. As assembly for refrigeration or generation will comprise pieces of p -type semiconductor and pieces of n -type, electrically in series with metallic connections between them, as in Figure 5. A theoretical study of the variables in Equation 1 shows that the semiconductor should be extrinsic up to the highest temperature at which it is to be operated, and that it should be 'doped' to give a thermoelectric power at its mean operating temperature near $+200 \mu\text{V per } ^\circ\text{C}$ for the

p -type, and near $-200 \mu\text{V per } ^\circ\text{C}$ for the n -type, so that the value for a single complete junction is near $400 \mu\text{V per } ^\circ\text{C}$. A single junction may be used for low powers, or an array as in Figure 5 for higher powers.

The requirements that the semiconductor should be extrinsic and not intrinsic at the operating temperature, and that α should be near $\pm 200 \mu\text{V per } ^\circ\text{C}$, set a lower limit to the forbidden gap of the material. This value of α determines the density n of free carriers, electrons or holes. This in turn for a given material determines the electrical conductivity

$$\sigma = ne\mu$$

where μ is the mobility of the charge carriers and e the charge on the electron. For many semiconductors the value of n is found to be of the order 10^{18} per cubic centimetre, representing an impurity concentration of the order 0.1 per cent. With this value of n , the minimum forbidden gap for operation at room temperature must exceed 0.1 eV, and for operation at 400°C , 0.25 eV.

Provided the conditions of the previous paragraph are met, the figure of merit is found to be proportional to

$$\frac{\mu(m^*)^{3/2} T^{3/2}}{K_0}$$

Here T is the absolute temperature, K_0 is the component of thermal conductivity due to the vibrations of the crystal lattice and m^* the 'density-of-states' effective mass of the charge carriers. The effective mass depends on the relationship between energy and momentum, which in solids is in general rather complex. It is determined by the

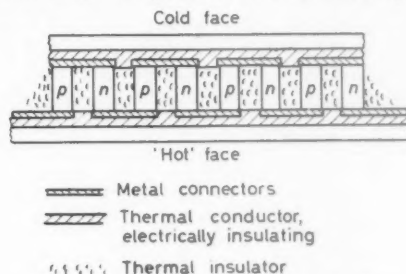


Figure 5. Array of thermojunctions electrically in series

energy band structure, which involves the shape and position of the energy-momentum surfaces in 'momentum space'. The density-of-states effective mass involves both the shape and the number N of such surfaces. The mobility itself depends on an effective mass parameter, but this is the 'inertial' mass m_i , involving only the shape of a particular

energy surface. The mobility is given by

$$\mu = e\tau/m_i$$

where τ is the time between successive collisions of the charge carrier with features of the crystal lattice which cause scattering. In the simple case of scattering by acoustic mode vibrations in a crystal with covalent bonding,

$$\tau \propto N(m^*T)^{-3/2}$$

The net effect of the other terms in the detailed expression for τ does not vary much comparing different semiconductors. Thus for this simple case the figure of merit becomes proportional to

$$N/m_i K_0$$

It is not so far possible to predict what m_i will be for a semiconductor from its atomic properties and structure. It is known that in general m_i falls as atomic weight rises and as energy gap falls. Similarly, though general features of the band structure can be deduced, the value of N cannot be predicted with certainty. There is no trend with atomic weight. Neither is the theory of thermal conductivity yet adequate to calculate the value of K_0 , though its dependence on the elastic properties enables some general observations to be made. Thus K_0 falls as atomic weight rises in a particular class of compound, and as ionicity in the bond

weight in this series. For these materials, however $N = 1$, and the values of K_0 are rather high, so that the figure of merit is poor. In these compounds there is some ionicity in the bond, *i.e.*, some charge on neighbouring atoms, so that the above theory does not apply exactly; however this does not affect the general argument. Germanium and silicon have respectively for electrons, $N = 4$ and $N = 6$, and the values of m_i are not very high. Again however, K_0 is too high to give a useful figure of merit.

The only group II to group VI compounds with high mobilities and low values of m_i are HgTe and HgSe. These also have low values of K_0 . The energy band structure has not yet been clarified, but in any case HgTe itself is of no value because of a very low energy gap, about 0.02 eV. Work is in progress on HgSe and alloys formed from it.

With any particular compound, replacement of one type of atom by another from the same column of the periodic table will not alter the doping level or the carrier density, but will lower K_0 by distorting the lattice. In some cases it will lower the mobility also, which is undesirable, but if it does not, then the figure of merit will be improved by such substitution.

Electrical and Thermal Properties of Bi_2Te_3

Infra-red measurements by I. G. AUSTIN⁹ in this laboratory have shown that Bi_2Te_3 is a semiconductor with a forbidden gap of near 0.13 eV at room temperature. When doped as an extrinsic semiconductor with a thermoelectric power of 200 μV per $^\circ\text{C}$, it has an electrical conductivity σ of about $10^3 \text{ ohm}^{-1} \text{ cm}^{-1}$, and a carrier concentration exceeding 10^{18} per cubic centimetre. At this level, it remains extrinsic up to about 200 $^\circ\text{C}$.

The thermal conductivity K for a solid is the sum of the lattice vibration component K_0 and the electronic component K_e , which is proportional to σ . A plot of K against σ for Bi_2Te_3 is shown¹⁰ in Figure 6. This is taken from the results of H. J. GOLDSMID, who has made a detailed study of the thermal conductivity and thermoelectric power of Bi_2Te_3 and its alloys. There is a sharp increase in K at low values of σ , which can be interpreted in terms of the transport of ionization energy as the material becomes intrinsic. The inhomogeneities in nearly compensated material may however play some part. Disregarding this effect, which is additional to K_0 and K_e , extrapolation of the plot to zero σ gives the value of K_0 . This is 0.0157 watt/cm $^\circ\text{C}$ for Bi_2Te_3 at 17 $^\circ\text{C}$ for heat flow in the direction of the cleavage planes. In the perpendicular direction it is about half as great. When σ is $10^3 \text{ ohm}^{-1} \text{ cm}^{-1}$, K_e is near 0.004 watt/cm $^\circ\text{C}$;

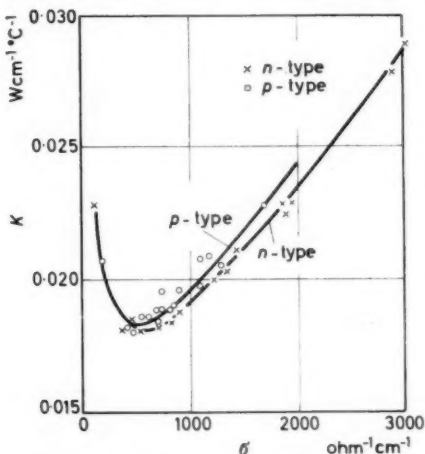


Figure 6. Relation between thermal and electrical conductivity for Bi_2Te_3

increases; it is lower the more distorted the crystal structure, and over a wide temperature range it is approximately proportional to $1/T$.

Particularly high mobilities and low values of m_i are encountered for electrons in some of the group III to group V compounds, *e.g.*, GaAs, InAs and especially InSb, with the highest atomic

combining these figures for α , σ , K_0 and K_e gives a figure of merit z between 2.0 and 2.2×10^{-3} at 17°C for samples cut so that the current flows along the direction of the cleavage planes.

The question arises whether the rather complex structure of bismuth telluride plays any important part in producing its high figure of merit—no other single compound can so far be made reliably and reproducibly with a figure of merit greater than 2.0×10^{-3} , though higher values can be obtained with several solid solutions. Although the value of K_0 is very low, several compounds are known with still lower values, so that this is not the only factor.

The crystal symmetry of Bi_2Te_3 indicates that in the band structure the number of valleys N is likely to be 1, 3, 6 or 12. The measurements of magnetoresistance by J. R. DRABBLE and his co-workers^{11,12} support the value 6 for both p - and n -type crystals, taking the most generalized interpretation. This implies valleys with their minima centred on the reflection planes. If however the valleys are in the most extreme position, at the edge of the Brillouin zone, their number would be 3 for the same scheme of interpretation. The preliminary results from Austin's work on Faraday rotation¹³ indicate that N is 3 for p -type crystals and 6 for n -type.

Knowing the shape of the energy-momentum surfaces, and knowing also that the electron and hole motion above 50°K (-223°C) is determined primarily by acoustic-mode lattice scattering^{12,14}, the Hall coefficient R can be used to determine the charge-carrier density n . Knowing the thermoelectric power α , the density of states effective mass m^* can be determined. Then from the value of N and the shape of the valleys, the individual mass parameters can be calculated. The drift mobility μ can also be determined; this is different from the 'Hall mobility' $R\sigma$. Present indications are that, assuming $N = 6$, the drift mobility for Bi_2Te_3 is $1250 \text{ cm}^2/\text{volt sec}$ for electrons and $515 \text{ cm}^2/\text{volt sec}$ for holes, and that the inertial effective masses for both are 0.09 times the free electron mass. The Hall mobility is near $400 \text{ cm}^2/\text{volt sec}$ for electrons and holes.

It may be noted that Bi_2Te_3 shows anomalies in the temperature variation of the Hall coefficient¹⁵ and the thermoelectric power¹⁶, both of which could be due to impurity bands overlapping the conduction and valence bands. The high dielectric constant, 85, would assist such overlap. The absorption edge shows a structure (Figure 7) interpreted by Austin⁹ as due to indirect transitions. While this is possible, the absorption coefficient is rather high, and it seems possible that both the

valence band and conduction band are double, with two maxima and two minima respectively, separated slightly in energy. Such a band structure would offer an alternative explanation for both the

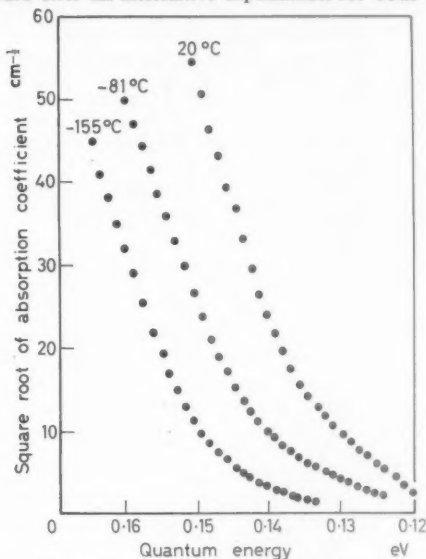


Figure 7. Infra-red absorption edge for Bi_2Te_3 .

absorption edge structure and the two anomalies referred to above. Such a structure has some support also from the two types of bonding in the crystal, described earlier. If this were the correct model, N would be doubled compared with the figures quoted, and the inertial mass would be smaller and the drift mobility larger.

It does appear therefore that the structure of bismuth telluride is one which can allow of a relatively small inertial mass—leading to relatively high mobility, and yet a fairly large density-of-states mass, because of a large value of N . The latter permits as required a high electrical conductivity for a given thermoelectric power.

Related Compounds

Antimony telluride Sb_2Te_3 has the same structure as Bi_2Te_3 . It has not been possible so far to prepare crystals with a carrier density much below 10^{20} per cubic centimetre, nor a thermoelectric power greater than $80 \mu\text{V per } ^\circ\text{C}$, so that it is doubtful whether it is a semiconductor. The lattice and electronic thermal conductivities have not so far been separated; the sum of the two is near $0.04 \text{ watt/cm } ^\circ\text{C}$, for σ about $4000 \text{ ohm}^{-1} \text{ cm}^{-1}$. The Hall mobility $R\sigma$ for holes is similar to that for Bi_2Te_3 .

energy surface. The mobility is given by

$$\mu = e\tau/m_i$$

where τ is the time between successive collisions of the charge carrier with features of the crystal lattice which cause scattering. In the simple case of scattering by acoustic mode vibrations in a crystal with covalent bonding,

$$\tau \propto N(m^*T)^{-3/2}$$

The net effect of the other terms in the detailed expression for τ does not vary much comparing different semiconductors. Thus for this simple case the figure of merit becomes proportional to

$$N/m_i K_0$$

It is not so far possible to predict what m_i will be for a semiconductor from its atomic properties and structure. It is known that in general m_i falls as atomic weight rises and as energy gap falls. Similarly, though general features of the band structure can be deduced, the value of N cannot be predicted with certainty. There is no trend with atomic weight. Neither is the theory of thermal conductivity yet adequate to calculate the value of K_0 , though its dependence on the elastic properties enables some general observations to be made. Thus K_0 falls as atomic weight rises in a particular class of compound, and as ionicity in the bond

weight in this series. For these materials, however $N = 1$, and the values of K_0 are rather high, so that the figure of merit is poor. In these compounds there is some ionicity in the bond, *i.e.*, some charge on neighbouring atoms, so that the above theory does not apply exactly; however this does not affect the general argument. Germanium and silicon have respectively for electrons, $N = 4$ and $N = 6$, and the values of m_i are not very high. Again however, K_0 is too high to give a useful figure of merit.

The only group II to group VI compounds with high mobilities and low values of m_i are HgTe and HgSe. These also have low values of K_0 . The energy band structure has not yet been clarified, but in any case HgTe itself is of no value because of a very low energy gap, about 0.02 eV. Work is in progress on HgSe and alloys formed from it.

With any particular compound, replacement of one type of atom by another from the same column of the periodic table will not alter the doping level or the carrier density, but will lower K_0 by distorting the lattice. In some cases it will lower the mobility also, which is undesirable, but if it does not, then the figure of merit will be improved by such substitution.

Electrical and Thermal Properties of Bi_2Te_3

Infra-red measurements by I. G. AUSTIN⁹ in this laboratory have shown that Bi_2Te_3 is a semiconductor with a forbidden gap of near 0.13 eV at room temperature. When doped as an extrinsic semiconductor with a thermoelectric power of 200 μV per $^\circ\text{C}$, it has an electrical conductivity σ of about $10^8 \text{ ohm}^{-1} \text{ cm}^{-1}$, and a carrier concentration exceeding 10^{18} per cubic centimetre. At this level, it remains extrinsic up to about 200 $^\circ\text{C}$.

The thermal conductivity K for a solid is the sum of the lattice vibration component K_0 and the electronic component K_e , which is proportional to σ . A plot of K against σ for Bi_2Te_3 is shown¹⁰ in Figure 6. This is taken from the results of H. J. GOLDSMID, who has made a detailed study of the thermal conductivity and thermoelectric power of Bi_2Te_3 and its alloys. There is a sharp increase in K at low values of σ , which can be interpreted in terms of the transport of ionization energy as the material becomes intrinsic. The inhomogeneities in nearly compensated material may however play some part. Disregarding this effect, which is additional to K_0 and K_e , extrapolation of the plot to zero σ gives the value of K_0 . This is 0.0157 watt/cm $^\circ\text{C}$ for Bi_2Te_3 at 17 $^\circ\text{C}$ for heat flow in the direction of the cleavage planes. In the perpendicular direction it is about half as great. When σ is $10^8 \text{ ohm}^{-1} \text{ cm}^{-1}$, K_e is near 0.004 watt/cm $^\circ\text{C}$;

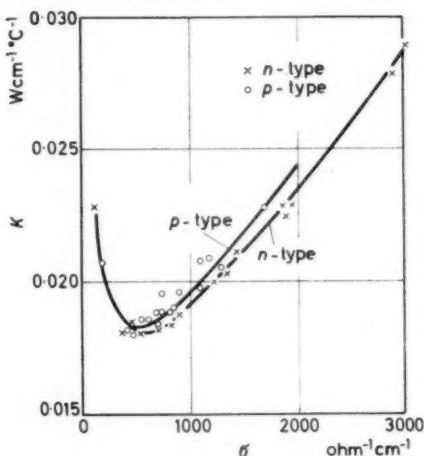


Figure 6. Relation between thermal and electrical conductivity for Bi_2Te_3

increases; it is lower the more distorted the crystal structure, and over a wide temperature range it is approximately proportional to $1/T$.

Particularly high mobilities and low values of m_i are encountered for electrons in some of the group III to group V compounds, *e.g.*, GaAs, InAs and especially InSb, with the highest atomic

combining these figures for α , σ , K_0 and K_e gives a figure of merit z between 2.0 and 2.2×10^{-3} at 17°C for samples cut so that the current flows along the direction of the cleavage planes.

The question arises whether the rather complex structure of bismuth telluride plays any important part in producing its high figure of merit—no other single compound can so far be made reliably and reproducibly with a figure of merit greater than 2.0×10^{-3} , though higher values can be obtained with several solid solutions. Although the value of K_0 is very low, several compounds are known with still lower values, so that this is not the only factor.

The crystal symmetry of Bi_2Te_3 indicates that in the band structure the number of valleys N is likely to be 1, 3, 6 or 12. The measurements of magnetoresistance by J. R. DRABBLE and his co-workers^{11,12} support the value 6 for both p - and n -type crystals, taking the most generalized interpretation. This implies valleys with their minima centred on the reflection planes. If however the valleys are in the most extreme position, at the edge of the Brillouin zone, their number would be 3 for the same scheme of interpretation. The preliminary results from Austin's work on Faraday rotation¹³ indicate that N is 3 for p -type crystals and 6 for n -type.

Knowing the shape of the energy-momentum surfaces, and knowing also that the electron and hole motion above 50°K (-223°C) is determined primarily by acoustic-mode lattice scattering^{12,14}, the Hall coefficient R can be used to determine the charge-carrier density n . Knowing the thermoelectric power α , the density of states effective mass m^* can be determined. Then from the value of N and the shape of the valleys, the individual mass parameters can be calculated. The drift mobility μ can also be determined; this is different from the 'Hall mobility' $R\sigma$. Present indications are that, assuming $N = 6$, the drift mobility for Bi_2Te_3 is $1250 \text{ cm}^2/\text{volt sec}$ for electrons and $515 \text{ cm}^2/\text{volt sec}$ for holes, and that the inertial effective masses for both are 0.09 times the free electron mass. The Hall mobility is near $400 \text{ cm}^2/\text{volt sec}$ for electrons and holes.

It may be noted that Bi_2Te_3 shows anomalies in the temperature variation of the Hall coefficient¹⁵ and the thermoelectric power¹⁶, both of which could be due to impurity bands overlapping the conduction and valence bands. The high dielectric constant, 85 , would assist such overlap. The absorption edge shows a structure (Figure 7) interpreted by Austin⁹ as due to indirect transitions. While this is possible, the absorption coefficient is rather high, and it seems possible that both the

valence band and conduction band are double, with two maxima and two minima respectively, separated slightly in energy. Such a band structure would offer an alternative explanation for both the

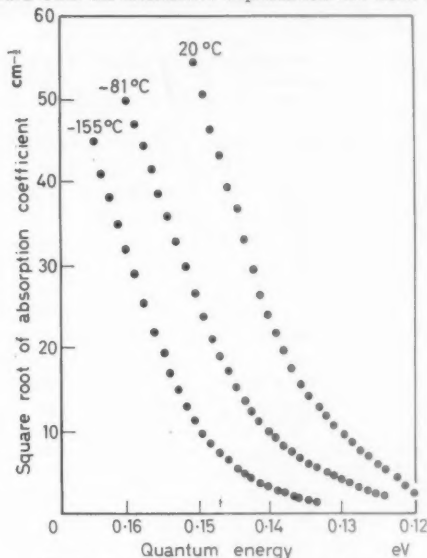


Figure 7. Infra-red absorption edge for Bi_2Te_3 .

absorption edge structure and the two anomalies referred to above. Such a structure has some support also from the two types of bonding in the crystal, described earlier. If this were the correct model, N would be doubled compared with the figures quoted, and the inertial mass would be smaller and the drift mobility larger.

It does appear therefore that the structure of bismuth telluride is one which can allow of a relatively small inertial mass—leading to relatively high mobility, and yet a fairly large density-of-states mass, because of a large value of N . The latter permits as required a high electrical conductivity for a given thermoelectric power.

Related Compounds

Antimony telluride Sb_2Te_3 has the same structure as Bi_2Te_3 . It has not been possible so far to prepare crystals with a carrier density much below 10^{20} per cubic centimetre, nor a thermoelectric power greater than $80 \mu\text{V per } ^\circ\text{C}$, so that it is doubtful whether it is a semiconductor. The lattice and electronic thermal conductivities have not so far been separated; the sum of the two is near $0.04 \text{ watt/cm } ^\circ\text{C}$, for σ about $4000 \text{ ohm}^{-1} \text{ cm}^{-1}$. The Hall mobility $R\sigma$ for holes is similar to that for Bi_2Te_3 .

Bismuth selenide Bi_2Se_3 is a semiconductor again with a similar structure and with a gap near 0.3 eV; Hall mobilities for electrons as high as 600 $\text{cm}^2/\text{volt sec}$ have been reported. The compounds Sb_2S_3 , Sb_2Se_3 and Bi_2S_3 are orthorhombic with larger energy gaps, exceeding one electron-volt, and smaller mobilities. The same applies to the monoclinic compounds As_2Se_3 and As_2Te_3 .

Alloys

A continuous series of solid solutions can be formed between Bi_2Te_3 and both Bi_2Se_3 and Sb_2Te_3 . In both cases the value of K_0 is reduced; it has a minimum value of 0.01 watt/ cm^2 °C near 50 atomic per cent for bismuth antimony telluride. The variation is more complex for Bi_2Te_3 - Bi_2Se_3 , and the minimum value is larger than in the system Bi_2Te_3 - Sb_2Te_3 .

Goldsmid's measurements indicate that $\mu m^{*3/2}$ for holes is fairly constant as antimony is substituted for bismuth, beyond 50 per cent antimony, and rises at higher percentages of antimony. Thus the figure of merit z for p -type material is higher for these alloys than for Bi_2Te_3 , a wide range of compositions giving values of z greater than 2.5×10^{-3} and some of them greater than 3.0×10^{-3} . Up to about 25 per cent antimony, the electron mobility is constant, but at higher antimony concentrations we have not found it possible to prepare n -type zone refined ingots with the required value of α of $-200 \mu\text{V per } ^\circ\text{C}$. Mobility of electrons has not therefore been studied beyond 25 per cent antimony. Near 20 per cent antimony, values of z for n -type material are near 2.7×10^{-3} . We have not so far been able to make optical studies of the energy gap in these alloys as the transmission has been very low. It is likely that there is a difference in detail in the band structure comparing Bi_2Te_3 and Sb_2Te_3 , and that in the alloys the change from one structure to the other occurs near 70 per cent antimony.

In Bi_2Te_3 - Bi_2Se_3 , the variation of gap width has been studied¹⁷, showing a discontinuity near 33 per cent selenium. We find the electron and hole mobility to be constant to about 10 per cent selenium, and then to fall; this does not agree with the Russian results on sintered alloys, for reasons which are not yet clear. The behaviour of the gap width indicates again a different band structure comparing Bi_2Te_3 and Bi_2Se_3 , and a change near 33 per cent selenium. In our experience so far these alloys give lower values of z than the bismuth antimony telluride series.

With a mean value of z for a junction of 3×10^{-3} , the maximum temperature difference obtainable with zero heat input to the cold junction is about 85°C. The coefficient of performance is 1.0 for

a temperature difference of 30°C, and 0.5 for 43°C. These represent the best figures for refrigeration obtainable so far, and are very encouraging, though inferior to the compressor type of refrigerator for powers exceeding a few tens of watts. As a thermoelectric generator, bismuth antimony telluride alloys are restricted because of their low energy gap; if intrinsic behaviour can be avoided up to 200°C in the best of these alloys, a thermojunction running between 200°C and 0°C would have an efficiency of nearly ten per cent.

All other compounds and solid solutions studied so far have values of z less than 3×10^{-3} at room temperature. They are correspondingly inferior to Bi_2Te_3 alloys for thermoelectric refrigeration, or for generation within the temperature range over which Bi_2Te_3 alloys can be used. The upper temperature limit of the latter is however near 200°C due to the onset of intrinsic behaviour, unless a considerable amount of selenium is added to raise the gap width. This however in our experience lowers the value of z . In any case operation much above 200°C is difficult because of atmospheric attack, and ultimately dissociation.

Materials suitable for operation above 300°C have so far values of z much less than 3×10^{-3} , in fact below 10^{-3} . The main question for the future of thermoelectric generation is whether materials can be found for operation at 400° to 1000°C, with values of z at least comparable with bismuth telluride alloys at 20°C, i.e., from 2 to 3×10^{-3} . Higher values than this seem very unlikely to be attained.

References

- GOLDSMID, H. J. and DOUGLAS, R. W. *Brit. J. appl. Phys.* (1954) **5**, 386
- WRIGHT, D. A. *Proceedings of 1956 Garmisch Conference, Halbleiter und Phosphoren*, p. 477 Braunschweig: Vieweg and Sohn, 1958
- *Proceedings of 1958 Brussels Conference* New York: Academic Press (In publication)
- GOLDSMID, H. J., SHEARD, A. R. and WRIGHT, D. A. *Brit. J. appl. Phys.* (1958) **9**, 365
- WRIGHT, D. A. *Nature, Lond.* (1958) **181**, 834
- FRANCOMBE, M. H. *Brit. J. appl. phys.* (1958) **9**, 415
- DRABBLE, J. R. and GOODMAN, C. H. L. *J. Phys. Chem. Solids* (1958) **5**, 142
- SHEARD, A. R. (In press)
- AUSTIN, I. G. *Proc. phys. Soc.* (1958) **72**, 545
- GOLDSMID, H. J. *ibid.* p. 17
- DRABBLE, J. R., GROVES, R. D. and WOLFE, R. *ibid.* (1958) **71**, 430
- DRABBLE, J. R. *ibid.* (1958) **72**, 380
- AUSTIN, I. G. *J. Electronics & Control* (1959) **6**, 271
- BOWLEY, A. E., DELVES, R. and GOLDSMID, H. J. *Proc. phys. Soc.* (1958) **72**, 401
- YATES, B. *J. Electronics & Control* (1959) **6**, 26
- GOLDSMID, H. J. *Proc. phys. Soc.* (1958) **61**, 633
- AUSTIN, I. G. and SHEARD, A. R. *J. Electronics & Control* (1957) **2**, 236

PRECISION ELECTRICAL MEASUREMENTS

L. HARTSHORN

Formerly of Electricity Division, The National Physical Laboratory, Teddington, Middlesex

This article on the techniques and problems of measuring electrical quantities with a very high degree of accuracy follows earlier articles on the precision measurement of time, length, temperature and heat capacity*. Further articles, on the measurement of radiant power and the use of statistical techniques applied to precision measurements, will be published at a later date.

ELECTRICAL methods of measurement are so extraordinarily convenient, sensitive and flexible, that they are nowadays used for the measurement of the quantities occurring in almost every branch of science and technology. Modern developments in measuring technique depend increasingly on circuit elements in which the change of some electrical property is accurately correlated with some other property, which thereby becomes measurable electrically. Indeed, the subject of precision electrical measurements is now so vast that whole volumes would be needed to do justice to all the recent developments.

To keep the subject within manageable bounds attention will be confined here to those electrical measurements that are considered to be basic to the whole scheme, so that they are performed in the national standardizing laboratories with the object of providing the standards and methods of measurement which serve as a reference for the electrical measuring techniques used elsewhere. Such measurements must obviously be made with the highest accuracy that is attainable.

Precision, Stability and Accuracy

The most accurate measurements in physical science today are the comparative measurements of frequency standards. Accuracies of one part in 10^{10} are often quoted in this field and measurements significant to a few parts in 10^{11} and even 10^{12} have been claimed. Such accuracies are only obtainable with electrical vibrators, such as quartz oscillators and atomic resonators, and the techniques involved are mainly electrical, so that it might be said that these are the most accurate electrical measurements made today. Moreover, since they serve to determine our standard scales of frequency and time, no measurement can be more fundamental. No purely electrical quantity can be measured with anything like the accuracy obtained with frequency; the highest accuracy obtained for the electrical quantities is at least ten thousand times worse and it is instructive to consider the reason for this.

The limiting factor is the constancy of the object to be measured, that is to say of the standard, since any physical object that can be successfully measured with the precision now under consideration will serve as a working standard for that quantity. The astonishing accuracy of the frequency standards has been obtained by searching for vibrating systems of higher and higher constancy, and indeed it can only be obtained for measurements on these highly selected systems and is not likely to have any bearing on general physical measurements until a comparable constancy has been achieved in other types of apparatus. If we consider the purely electrical quantities—resistance, current, voltage, capacitance, inductance, etc.—and the corresponding standards, we find that even the best standard resistors, cells, capacitors and inductors are very rarely constant to one part per million over a few months and therefore, although it is possible to make comparisons of these quantities on a particular occasion to say one part in 10^6 , there is little point in the exercise because the standards and the apparatus employed may change by relatively large and unknown amounts by the time they are next used.

The effective accuracy of the measurement is therefore limited by the secular variations of the standards. These variations must be observed by making measurements at intervals over a long period and such measurements together with continued efforts to devise standards of better constancy, or 'stability', form an essential part of the work of the standardizing laboratories. (The latest results in this field are always recorded by the international committee¹ concerned with these matters.)

In considering electrical standards, it will be useful to begin with standards of resistance and standard cells since these are indispensable to every electrical laboratory. Standards of capacitance follow as next in order of general utility and then those of inductance, which although perhaps less convenient than the others for general measurements are important as the starting point of absolute measurements, which will be discussed later.

* *Research, Lond.* (1957) 10, 198, 217; (1958) 11, 147, 192

Resistance Standards

Probably the best performance so far recorded for resistance standards is that shown in *Figure 1* for a group of five coils, each of nominal value one ohm, which are used at the National Physical Laboratory as the reference standard of resistance. The radius of each of the small circles marking a measured value corresponds to a resistance deviation of 2×10^{-7} ohm and it is evident from the general consistency of the results over a period of twenty years that the precision of the comparisons is not worse than about 1×10^{-7} ohm. The coils do, however, show an appreciable drift in value relative to one another, the greatest deviation amounting to 8×10^{-7} ohm in twenty years. It has been assumed that the mean value of these coils, which have been specially selected, remains constant since there is nothing to suggest that any one is better than another; ordinary commercially made standards are likely to show considerably greater changes. It is of course probable that this mean value is drifting, for past records have shown that groups of similar coils drifted together at a rate that was at least as great as the relative drift between any two members.

The best available evidence for the constancy of the mean value is that given by comparisons made, from time to time, between the coils shown in *Figure 1* and the standards maintained at the National Bureau of Standards, Washington and the International Bureau of Weights and Measures at Sèvres. These were reported² as amounting to a deviation of about 2×10^{-6} ohm in ten years between the mean values of these three groups of coils. Since the relative constancy of the coils of the N.B.S. group is of the same order as that of the N.P.L. group, it is probable that the mean value of each group drifts at a rate at least as great as any of the relative deviations within the groups.

The factors which make for high stability in standard resistance coils have been studied at the various national laboratories and of all the resistance alloys examined manganin has so far given the best results. Both the N.P.L. and N.B.S. coils are of annealed manganin, the former in a container filled with a pure paraffin oil and the latter in air.

Reduction of temperature coefficient of resistance is much more important in day to day precision measurements than in measurements on reference standards where temperature is always finely controlled. At the N.P.L. a nickel-chromium-aluminium alloy has been found very suitable for working resistance standards since its temperature coefficient can be adjusted by heat treatment and it possesses a higher resistivity than either manganin or constantan. This property is important in precision

a.c. work since, other things being equal, the higher the resistivity of the material, the smaller the linear dimensions of the resistor and therefore the smaller the unavoidable but unwanted residual inductance and capacitance and the errors in phase-angle associated with them.

Standard Cells

The practical standard of voltage or potential difference is the standard cell and each of the national standardizing laboratories employs a selected group of cells as the primary reference standard by means of which the theoretical unit, the volt, is realized and preserved. The cells now used are all of the Weston type, the electrodes being pure mercury and pure cadmium amalgam and the electrolyte a solution of cadmium sulphate. In the construction of reference standards stability is the main criterion and experience has shown that the best performance is given by cells in which the electrolyte contains excess cadmium sulphate crystals and the mercury electrode is covered with a depolarizing layer which consists of a paste of mercurous sulphate, cadmium sulphate and mercury. The actual value of the e.m.f. of the cells is not of the first importance; it is settled in practice, not by specification but by international agreement based on absolute considerations following international comparisons of the reference groups. Such comparisons are necessarily limited in precision by the stability of the group of reference cells.

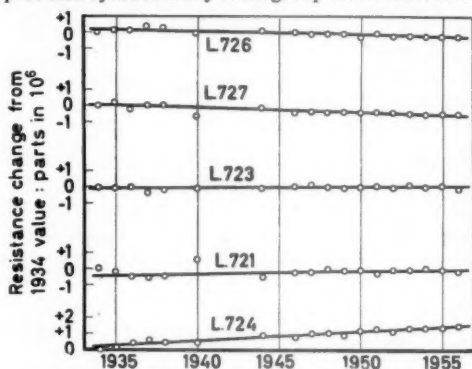


Figure 1. Observed changes in the reference group of one ohm resistors at the N.P.L.

It is scarcely to be expected that a chemically active system like a standard cell would show a stability comparable with that of an inert object like a resistance coil and, indeed, when a batch of cells has been made it may be necessary to discard one here and there because it shows unmistakable variations relative to the rest. As in the case of resistors, relative stability within a group is the only

evidence available but the cell has, of course, a more limited life and one that is ten to twenty years old is usually discarded. Thus a reference group may be changed with the years and the group will usually include cells of different ages. The volt is preserved by assuming that the mean value of the group remains constant so long as the deviations of members of the group relative to their mean value remain within tolerable limits; experiment alone can fix these limits.

An examination of the comparisons made at the N.P.L. during the last twenty years has shown that selected small groups of cells constructed between 1928 and 1937 have remained constant relative to the main group apart from apparently random variations, which were usually about $\pm 1 \times 10^{-6}$ volts and on no occasion exceeded 4×10^{-6} volts in this period. Comparisons between these cells and the reference group at the International Bureau suggest that their mean values have drifted apart by 2×10^{-6} volts in about sixteen years. Thus every measurement of voltage and of quantities dependent on voltage like current, power and energy is limited by unknown variations of at least this order for these are selected cells used only in specially favourable conditions, for example they are maintained at a constant temperature for a matter of weeks before any measurement is made.

Standards of Capacitance

The ordinary standard of capacitance may usefully be considered here because it is so widely used in a.c. measuring techniques. For values of capacitance exceeding about $0.01 \mu\text{F}$ mica capacitors are indispensable; for smaller values air capacitors are usually better, not only because, as a dielectric, air is more nearly perfect than mica, but also because the air capacitor being a more rigid structure is capable of better stability.

The difficulties encountered with mica capacitors arise partly from mechanical imperfections of the structure and partly from the fact that mica has a very strong tendency to take up water from the atmosphere which dissipates power and increases the capacitance by an amount which depends on the frequency. This is countered to some extent by clamping the mica between metal foil electrodes or by depositing metal films on the mica surface and sealing the assembly to prevent the intake of moisture but the performance of a capacitor cannot be predicted from a specification of its construction alone. Variations of capacitance with temperature are not always strictly reversible and there may also be small seasonal changes connected with imperfection of the sealing and changes in humidity, as has been shown by G. H. RAYNER and L. H. FORD².

With standards of good quality an accuracy of 5×10^{-6} is practicable but only in exceptional circumstances can an accuracy better than this be maintained with a mica standard.



Figure 2. The Campbell primary standard of mutual inductance of the N.P.L. (Crown Copyright Reserved)

The mechanical stability of air capacitors is very much better although the same difficulties appear to a minor extent in the solid insulators and in any surface film that may form on the metal surfaces. Such difficulties are largely overcome by using fused quartz or good ceramic insulators and clean surfaces and the only appreciable changes are those arising from variations of electrode spacing caused by the mechanical yielding of the structure with changes of stress. In order to obtain reasonably large capacitances, electrodes consisting of large banks of plates interleaved to give a small electrode spacing are used. The internal stresses in the banks of plates may be considerable, depending on the method and conditions of assembly and these stresses may change considerably with changes in temperature.

If the structure is satisfactory the temperature coefficient of capacitance will be equal to the coefficient of linear expansion of the metal but in many cases much more erratic changes are observed and the stability is correspondingly poor. It is almost impossible to quote a representative value except that an accuracy of $1 \times 10^{-4} \mu\text{F}$ is as much as can be expected of a capacitor that is moved about from one place to another. However, W. K. CLOTHIER³ at the Australian National Standards Laboratory has recently designed some fixed capacitors in which each plate system is supported by a single pillar, instead of the usual three, so as to secure freedom from internal stress and the first reports of the stability suggest that it is of an exceptionally high order, 1×10^{-6} being mentioned.

If a stability of this order proves to be reproducible, the standard air capacitor would become very important.

What has been said above applies equally to the variable air capacitor, but in this case there may also be errors arising from mechanical imperfections of the bearing and scale. These are usually checked by the observer by repeating readings and testing for backlash when making any observation requiring full accuracy, so that they are much better known than the more elusive errors mentioned above.

Primary Standards of Inductance

The system of electrical measurement that is practised today is an absolute system, that is to say the only standards that have a definite status are those of length, mass and time and all electrical techniques proper must begin with some electrical quantity that can be derived from measurements restricted to length, mass and time alone. Thus the first step is necessarily the establishment of some practical link between an electrical quantity and these basic mechanical quantities, which are in principle, sufficient to determine the whole electrical system. The links that have actually been constructed at the various standardizing laboratories fall into three groups: current balances or dynamometers giving current in terms of mechanical measurements; inductors giving an inductance in terms of length; and generators like the Lorenz machine, or forms of induction balance, that give resistance in terms of length and rotational speed or frequency.

Mutual inductance is a common element in all these links, and it follows that the accuracy of the whole system turns very largely on the accuracy with which inductors of suitable size and form can be constructed and evaluated from length measurements. These special inductors are often called primary standards of inductance. They do in practice serve as standards, the ordinary working standards being compared with them by electrical techniques, but in principle the only standard recognized is that of length, and any variation of the primary standard of inductance is theoretically of no importance since it can be detected and evaluated by repeating the measurements of length. Indeed, *in principle*, the absolute system overcomes all the difficulties associated with the changes in the electrical standard discussed above, but in practice the repetition of a set of mechanical measurements on a primary standard with the necessary accuracy may take months of arduous labour and constancy during this period and the subsequent period of use in electrical measurements is also a vitally important factor. It provides problems no less complex than those encountered with the other standards.

The primary standard of mutual inductance at the N.P.L. may be taken as typical of present day practice in this field. The original model was devised by Albert Campbell almost 50 years ago and although many improvements have been made in the details, the essential features still stand (Figure 2). The primary coil consists of a single

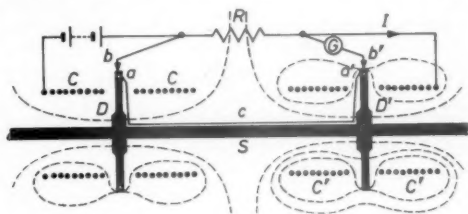


Figure 3. Simplified circuit of Lorenz apparatus illustrating Campbell's principle of inductor construction

layer solenoid wound in two sections on a non-magnetic insulating cylinder, some 30 cm in diameter, and this is encircled by a circular secondary coil, of about 500 turns but of small cross-section, placed symmetrically with respect to the two halves of the primary coil. The siting of the secondary coil with respect to the primary is important. Campbell realized that by winding the primary coil in two sections, suitably separated, the magnetic field obtained when current flows through the two sections in series includes a so-called neutral ring—a circle having zero magnetic field strength at every point on its circumference—midway between the two halves of the primary coil, and that if a circular secondary coil is made to occupy the position of this ring the mutual inductance between primary and secondary is approximately independent of small displacements and deformation of the secondary coil.

The idea is illustrated by Figure 3 which shows schematically the application of Campbell's principle to the Lorenz machine for the absolute determination of the ohm. Here the secondary coil is replaced by a rotating disc and it is obvious that with this configuration, where the rim of the disc occupies the neutral ring, that the magnetic flux threading the disc and therefore the mutual inductance is independent of small displacements or distortions of the disc. It follows that the accuracy of a Campbell inductor is almost wholly dependent on the accuracy with which the linear dimensions of the primary coil alone can be measured.

To ensure the highest accuracy such coils are made in the form of single layer solenoids wound with thin bare wire, usually of hard drawn copper under tension, wound on to a cylinder of insulating

material that has a low coefficient of expansion and great rigidity; the tension ensures that the coil dimensions are controlled by those of the cylinder. Accuracy in both the radial and axial dimensions of the coils is essential and thus every turn of the primary coil must be precisely located. This is done by winding the wire in a helical groove accurately cut into the surface of the cylinder. Marble, which has a coefficient of expansion of 5×10^{-6} per $^{\circ}\text{C}$ and tolerably good insulating properties, was originally used and is still giving good service. Better insulators are now available: glass and porcelain and fused quartz have all been tried. The latest coil to be constructed at the N.P.L. is of fused quartz and marks an appreciable step forward, but difficulties encountered when constructing a cylinder of fused quartz proved quite formidable.

A standard inductance must also be usable in a practical electrical circuit with at least the same precision as achieved in the mechanical evaluation. This means that the stray inductance associated with the connecting leads must be made negligibly small. The finished coil must have a high degree of uniformity for only then can the geometry of the whole circuit be inferred with the necessary accuracy and the inductance of the actual circuit, and not merely that of the ideal circuit to which it approximates, calculated. The helical groove be uniform in pitch, diameter and section, and the wire itself must be of uniform and circular cross-section and precisely located by line contact along the two sides of the groove. One of the early coils showed systematic irregularities in measured dimensions because the wire, though uniform, was slightly elliptical in cross-section and the ellipse had rotated during the winding.

The secondary coil need only be known to a much lower order of accuracy; the diameter of each layer is measured as it is wound and thus the mean diameter determined and also the overall dimensions of the cross-section of the winding. The spacing of the turns is adjusted, for example, by thin paper or mica between layers, so that the central turn is equal in diameter to the neutral ring; the main calculation is made for this turn and the values for the other turns obtained by applying suitable corrections. As already explained the first order term in these corrections is zero and the second order terms are obtainable with sufficient accuracy from coil dimensions of relatively low accuracy. Some 500 turns are required on the secondary coil to obtain a mutual inductance of about 10 mH which is a convenient value to handle electrically. Finally thin insulated wire is used and the turns are closely packed and the cross-section of the winding made nearly square so that no turn is farther than is

necessary from the neutral ring. The overall correction for a winding of this kind can be simply calculated from the dimensions of the cross-section alone and it proves to be remarkably small.

Having constructed and measured the two coils, made all the calculations and mounted the secondary coil in its correct symmetrical position by means of non-magnetic and non-conducting adjusting screws, the result is a standard of mutual inductance the value of which is known to a few parts per million. Such a standard is extremely limited in its application and the next step is to produce a good working standard with a wide range of application and to calibrate it by reference to the primary standard.

Working Standards of Inductance

The primary standard is essentially the concrete realization of one of the units of the absolute system, the centihenry in the example given. The standardizing laboratories must, however, not merely produce primary units but establish a system of measurements whereby quantities required in practice can be determined in terms of this unit. The henry is mainly required for a.c. measurements and a comprehensive system of such measurements has been built up at the N.P.L. The accuracy obtained varies with the nature and magnitude of the quantity to be measured and the frequency but obviously it cannot exceed, and seldom equals, that of the primary standard.

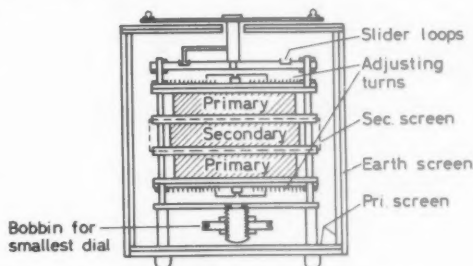


Figure 4. The N.P.L. working standard of mutual inductance

Although standards of self-inductance and of capacitance have recently become more widely used in industrial laboratories than standards of mutual inductance, largely because their values hold good over a wider range of frequencies, the principal working standard of the N.P.L. has remained a variable standard of mutual inductance. A standard of mutual inductance has practical advantages at lower frequencies and work of high accuracy is carried out at low frequencies. If a single primary coil and a number of secondary

coils connected in series are used—or what amounts to the same thing, a secondary coil subdivided into several parts which can be thrown into a circuit as required by means of a brush contact—then at low frequencies the mutual inductances associated with the primary and the parts of the secondary coil are strictly additive. (Due regard must be paid to the sign of the inductance.) Thus if the parts are separately calibrated by reference to a fixed standard, the series-connected system constitutes a standard scale of mutual inductance. The algebraic character of the addition is a great advantage because it makes possible the realization of a true zero of mutual inductance. Another advantage is the ease of realizing a continuous variation of mutual inductance; one has merely to move the secondary coil continuously relative to the primary and it is easy to find positions where part of the coil contributes mutual inductance of opposite sign to that of the remainder, then by continuous adjustment the magnitude of the two parts can be varied so that the net mutual inductance passes from positive values through zero to negative values. It was for this reason that Campbell devised as his working standard what he called an inductometer, a direct-reading continuously variable standard of mutual inductance providing every value between zero and whatever full scale value is convenient, 10 mH in the first instance.

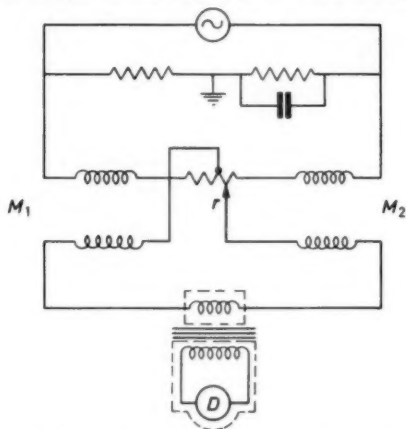


Figure 5. The mutual inductance comparator

The inductometer shown in Figure 4 is used at the N.P.L. The coils are multi-layer windings on marble bobbins, the primary and secondary coils are coupled as closely as possible and the total number of turns in each is of the same order so that the self-inductance of neither winding has to be made unduly large to obtain the required mutual inductance; a condition important in minimizing

the unwanted but inevitable effects of capacitance in the windings. In winding the secondary coil the ends of successive groups of turns estimated to give tenths, hundredths and thousandths of the total mutual inductance are connected to the contact

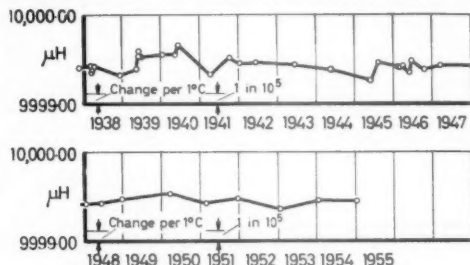


Figure 6. Observed changes in the N.P.L. working standard of inductance—these changes set the limits to the accuracy of many day to day measurements of electrical quantities

studs of three decade dials by means of which any desired value of mutual inductance is obtained to the nearest thousandth. Finer adjustments are obtained from a single circular turn at the end of the secondary winding; a radial conductor terminating in a brush contact sweeps round this circular turn and in effect provides a small secondary coil of sector shape, the area of which increases linearly with the angular motion of the conductor, which carries a pointer and thus gives a linear scale of mutual inductance. The scale extends over 360° and thus can be read to about one part in a thousand of the full scale reading or one part in a million of the total value of the standard. Adjustment and calibration of the various sections of the secondary is a matter of trial and error; due to the additive property, adjustment of any section is easily done by including in it an additional small loop the area of which is adjusted by squeezing here and there or turning it over if a negative change is needed.

The circuit used for such work and for subsequent measurements of mutual inductance is shown in Figure 5, which represents the simplest of all precision a.c. measurements. Two mutual inductors are balanced by connecting their primary coils in series with a generator and their secondary coils in series-opposition with a null-detector, which indicates equality of the two mutual inductances by zero response. For high precision the generator must supply a.c. of a frequency of 100 ~ or lower and of approximately sine wave form and the detector must be tuned to the same frequency.

It is then observed that the two opposed inductors alone never give a perfect balance but only a mini-

mum response which becomes sharper if the frequency is lowered. This difficulty arises because the e.m.f.'s induced in the two secondary coils are not exactly opposite in phase. Ideally, both e.m.f.'s are strictly in quadrature with the common primary current and should therefore balance exactly, but in practice the current in the primary circuit is not wholly confined to the two windings; some of it flows as displacement current through the capacitance between neighbouring turns of the windings and the two inducting currents are therefore not quite identical unless the capacitance, resistance and inductance distributions in the two inductors are exactly the same, which is not generally the case. In addition there are similar capacitances in the secondary windings and displacement currents will exist in these even when the null-detector indicates zero. Thus what is actually balanced is the secondary p.d., not in the ideal condition of zero secondary current but with the finite secondary currents determined by these capacitances. Another factor modifying the actual secondary p.d. is the existence of eddy currents within the metal of the windings and terminals.

All these extraneous factors diminish as the frequency is lowered but even at 50 c/s they are appreciable in work of high precision. A true balance becomes possible in spite of such factors if a small resistive potentiometer is included in the circuit as shown in Figure 5; a small adjustable p.d. in phase with the primary current is thereby injected

into the secondary circuit and if this potentiometer has a centre zero, as shown, this injected p.d. is easily adjusted to have either sign or to be strictly zero as may be necessary. Thus a unique balance point can always be obtained for any one connection of the coils and the reading of the potentiometer is a measure of the residual effects arising from capacitance and eddy currents.

This arrangement is used for the calibration of the working standard by reference to the primary standard, and since the calculated value of the primary standard can only be realized when all such effects are negligible, the frequency for this calibration must be lowered to a value at which the effects do prove to be negligible, *i.e.*, when a further lowering of the frequency causes no appreciable change in the balance point. With the Campbell standard a frequency as low as 10 c/s is found to be necessary for the precision required. In an ideal arrangement, reversal of any two coils would have no effect on the balance point but in practice a small change is usually observed, indicating the existence of finite cross mutual inductance between the primary of one inductor and the secondary of the other. It is easy to show that because of the additive property these stray mutual inductances are eliminated if the mean of the four readings obtained by making all the possible reversals of pairs of coils is taken. It is of course also necessary to short circuit the leads at the points where they enter the secondary coil of the primary standard and to verify

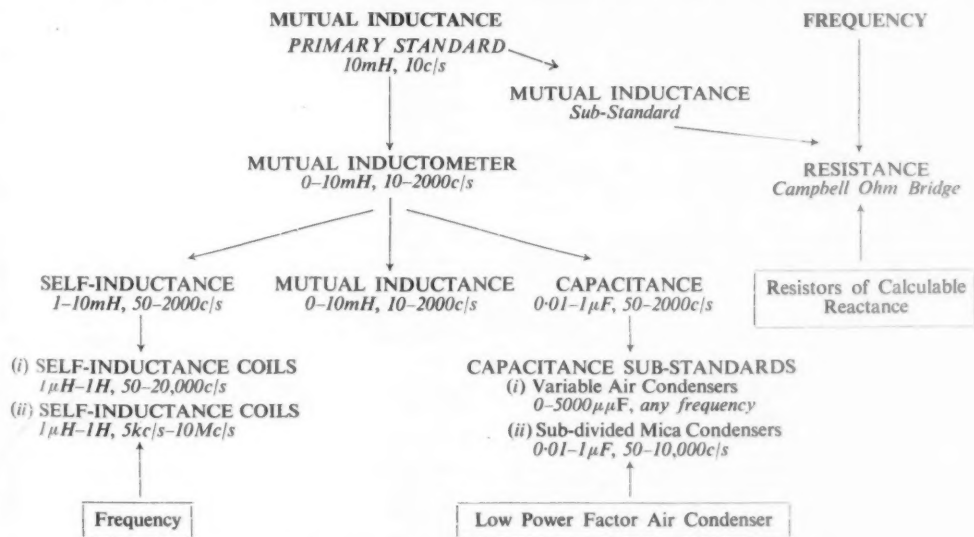


Figure 7. Part of the N.P.L. scheme of calibrations of working standards

that they contribute zero mutual inductance; coaxial cable like those seen in *Figure 2* or closely twisted bifilar leads to each coil are used.

The value of the working standard at the setting 10 mH is obtained by the above procedure. The sub-divisions of the working standard are then checked by comparison with one another by means of the same circuit but with the primary standard replaced by another inductor which merely serves as a constant ballast with which any two nominally equal sections of the working standard can be compared. Several procedures differing only in detail are possible but the principle is simply that used for calibrating a box of weights in terms of the largest or for calibrating the sub-divisions of a scale of any other additive quantity. The circuit is essentially a comparator of a sensitivity which can be made almost as high as we please by incorporating a suitable amplifier in the detector. It is not difficult to make observations with a precision of one part per million but the instability of the standards and uncertainties in the realization of the absolute unit limit the accuracy obtainable at any given time to a somewhat lower level.

The constancy of the primary standard depends almost entirely on the dimensional stability of the primary cylinder which can be checked from time to time by length of measurements, but the few observations on record make it difficult to say more than that a marble standard may show irregular changes but they are probably within ± 5 parts per million in ten or twenty years, and there is reason to believe that the new fused quartz standard will be much better.

The working standard is obviously a much less stable structure and its variations are recorded by making regular comparisons with the primary standard. Campbell's original inductometer changed by a few parts in 10^4 in the course of ten years or so. The coils had been impregnated with paraffin wax to prevent the entry of atmospheric moisture into the silk covering of the wire. The wax itself, however, caused some instability probably because of stresses set up in the windings as it solidifies. The coils of the present standard were impregnated with oil only and the whole instrument was filled with a semi-fluid mixture of transformer oil and petroleum jelly. The order of constancy achieved with this construction is shown in *Figure 6*. When due allowance is made for the temperature coefficient, which is itself $1.2 \times 10^{-5} \text{ deg C}^{-1}$ the deviation from the mean value has not exceeded $\pm \frac{1}{2}$ part per 10^5 in ten years. This sets the limit to which every day measurements of capacitance and inductance can be made. So far as is known no working standard has given a better performance.

Further Developments

It is impracticable to pursue in detail the whole scheme of measurements by which all the electrical magnitudes are realized in accordance with the internationally accepted absolute system, but the basic steps described above give some idea of the nature of the researches now being pursued at the standardizing laboratories. An idea of the extent of the work can perhaps be best conveyed by the diagram of *Figure 7* which shows the part of the scheme of development followed at the N.P.L. that bears directly on the operations discussed, and by mentioning some of the major researches reported as in progress at other national laboratories.

At the National Bureau of Standards, Washington, a new absolute determination of the ampère is being made by means of a dynamometer of the Pellat type. This will supplement an ampère balance of the Rayleigh type which has been brought to a remarkably high state of development in recent years. Another current balance on similar lines was established at the Institute of Metrology of the U.S.S.R. in 1955. At the German National Laboratory (D.A.M.G.) the use of a calorimetric method for the determination of the ampère is under investigation. The heat is generated by arresting the motion of freely-falling weights within the calorimeter, the fall being made repetitive by rotating the calorimeter. This is a new principle in the field of absolute measurements and therefore of special interest, although the difficulties are obviously very great.

Another break with tradition is being made at the National Standards Laboratory of Australia where an absolute standard of capacitance is being developed as an alternative to the primary standard of inductance described above. This is equivalent to the use of an electrostatic system of measurement instead of the electromagnetic system usually followed. However, the velocity of light which determines the conversion factors for the two systems is now known with sufficient accuracy to make the investigation a new approach to the accepted system rather than an experiment in a new system. Many other investigations could be mentioned but enough had probably been said to show how the precision nowadays taken for granted in electrical measurements had been developed and is being continually improved.

References

- ¹ *Procès-verbaux du Comité International des Poids et Mesures*, Vol. 26B (in the press)
- ² RAYNER, G. H. and FORD, L. H. *J. sci. Instrum.* (1954) **31**, 3
- ³ CLOTHIER, W. K. *Proc. Instn elect. Engrs* (1954) **101**, Part II, 453

SOLID POLYURETHANE ELASTOMERS

S. H. MORRELL

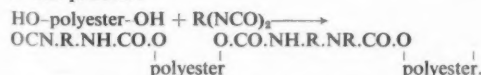
Research Association of British Rubber Manufacturers, Shawbury, Shrewsbury

A review of non-cellular polyurethanes is given which includes the results of an evaluation of some typical rubbers by the Research Association of British Rubber Manufacturers. Applications are discussed and it is concluded that although casting elastomers possess the best physical properties, the processable types may be preferable where the highest physical properties are not required.

AMONG the wide range of synthetic rubbers now available the polyurethanes are in many respects a unique group. Their chemistry is so unusual compared with that of other rubbers, and their properties and applications are so diverse, that a special study is desirable by all who wish to keep abreast of modern developments. The basic chemical reactions by which polyurethane elastomers are prepared may be varied to give different end products. It is thus possible to obtain both non-cellular and cellular materials, and furthermore the non-cellular group may be further sub-divided into those elastomers which can be processed on conventional rubber machinery and those which have to be given their final shape by means of a casting process. It is proposed to describe the non-cellular materials in the present article: the cellular materials will form the subject of a further paper.

Chemistry of a Typical Casting Rubber

Waxy polyesters of molecular weight 2000 to 5000, formed for example by the reaction between ethylene glycol and adipic acid, are reacted with an excess of diisocyanate: this increases the molecular weight to between 4000 and 15,000, and also provides reaction centres which are utilized in the next stage of the process:

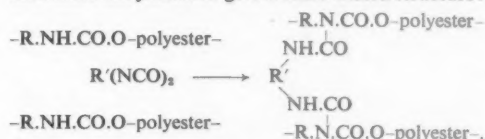


The chains are then further extended by the action of glycols or diamines:



The mixture is cast in open moulds, when reaction occurs between the urethane linkages and

the excess *isocyanate* to give a cross-linked structure:



The reaction is completed by heating the mould for a short time: the casting is then removed and the finished product given an additional oven cure.

Technology of Casting Polymers

A brief account of the manufacture of a typical casting polymer is not out of place, as it is a process unfamiliar to the majority of rubber manufacturers.

A mixing vessel is required, made of mild steel or stainless steel, in which the liquids may be stirred under reduced pressure. Means of heating and cooling the contents should also be provided, and the whole interior must be capable of easy cleaning. Since the polyester is hygroscopic, it must be dehydrated by heating and stirring at a temperature slightly above 100°C, whilst the pressure is reduced to 15 to 20 mm of mercury. The liquid should then be cooled before the *isocyanate* is added. The pressure is again reduced whilst the liquid is stirred for a few minutes to remove any carbon dioxide formed by the traces of water still remaining. The liquid is then discharged into warm moulds, which have previously been coated with a suitable mould release to prevent adhesion of the polyurethane. The cure is completed by heating above 100°C for a few hours. (A typical apparatus is shown in *Figure 1*.)

In an alternative process, polyethylene adipate is reacted with naphthalene 1:5-diisocyanate, and this is followed by cross-linking with butane-1:4-diol. The polyester is first dehydrated as before

and stored in a molten state until required for use. It is then mixed with the *isocyanate* in a stainless steel steam-jacketed autoclave under reduced pressure, and the reaction allowed to proceed until the peak of the exothermic reaction has been



Figure 1. Apparatus for making solid polyurethane casting polymers (by courtesy of Dunlop Rubber Co. Ltd)

reached, when the diol is added. The resulting syrup, at a temperature of about 120°C, is discharged into wax-coated moulds. Set-up occurs rapidly and the product can be removed from the mould after 20 minutes: an oven cure of 24 hours at 110°C is then required to complete the process.

Technology and Chemistry of Processable Polymers

The processable polymers are supplied as rubber-like material which may be compounded with other ingredients on two-roll mills or in internal mixers of the Banbury or Intermix type. The stock may then be processed further by extrusion or calendering before curing in a press, or in hot air. Curing in steam is not advisable because hydrolysis of the urethane linkages takes place.

The chemistry of the processable polymers has not been widely published. The same general reaction scheme holds as for the casting rubbers, but the reactions have to be carefully controlled so

that a rubber-like state is reached in which there are no free *isocyanate* groups to cause cross-linking during processing and storage. In the earlier polymers of this kind, curing was effected by means of additional *isocyanate*, but more recently peroxide and accelerated sulphur curing has become possible. These modifications are discussed further in the next section.

Some Proprietary Polyurethanes

In this section will be given descriptions of some typical polyurethanes, particularly those whose physical properties are described later. In some cases, however, firms are manufacturing under licence from the original patentees polyurethanes bearing their own trade names; such polymers may contain special additives to produce certain desirable properties. When this is known to have occurred, only a brief mention is made of the rubber. It must not be assumed, however, that the properties of such compounds will resemble in all respects those of the polymers produced by the original patentees.

Vulkollan, the earliest polyurethane elastomer, was first produced in Germany during the war, and its preparation follows closely the scheme given above. (*Vulkollan* gear wheels are illustrated in Figure 2.)

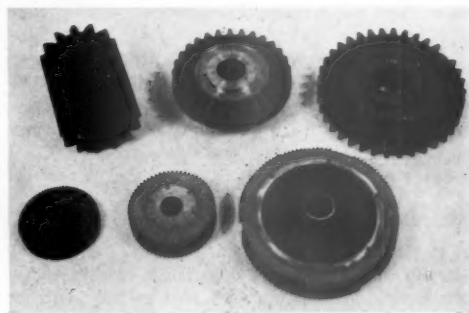


Figure 2. Gear wheels of various types made from *Vulkollan* (by courtesy of Farbenfabriken Bayer A.G.)

Duthane, *Prescollan* and *Ulon* are similar materials made under licence in this country, whilst there are also several available in America under various trade names. In all cases the polymer has to be prepared as required since its storage life is short, and it is only possible to prepare products by casting without pressure the fluid intermediate at a temperature of 100° to 125°C. Loading with reinforcing fillers is not effective since it is difficult to produce reinforcement where milling of the solid elastomer is not involved. The ingredients for

making polyurethane rubbers of this molecular type are also available in Germany and the United States. In Germany the polyester is marketed under the trade name *Desmophen* and the isocyanates under the name *Desmodur*, whilst in the United States the names *Multron* and *Mondur* respectively are used. In the United Kingdom, ingredients are manufactured for preparing soft rubbers of hardness 15 to 35 BS°: the polyester is called *Dalterol PR 1* and the isocyanate *Suprasec SF*.

Parallel work in the United Kingdom has been conducted with polyesteramides, formed by the condensation of ethylene glycol, adipic acid, and ethanalamine, followed by treatment with xylene to remove the water. From this emerged, in 1943, *Vulcaprene*, an elastomer which could be worked on normal rubber machinery but which had inferior physical properties to the Vulkollan type. It has, however, wide use in lacquers and in protective coatings for fabrics, but is rarely used in the solid form on account of its low tensile strength. A comparison of Vulkollan N and Vulcaprene A made by the Research Association of British Rubber Manufacturers in 1947 is described later.

Chemigum SL was introduced in the United States in 1953 as an elastomer processable on normal rubber machinery. It appears that the fundamental polyurethane chain of Vulkollan is modified but slightly, and that an improved storage life of some weeks is attained by careful control of the number of chemically active end-groups, and also of the amount of chain extension effected with diisocyanate. Special additions are used to eliminate 'blowing'

in the final vulcanizing stage of the reaction, a phenomenon which results in porous products. Black loading may be employed up to ten or fifteen parts per hundred of rubber, with beneficial effect on tear strength, but the addition of further amounts causes a deterioration in physical properties. Unfortunately building tack is poor, but may be improved by the judicious use of a solvent.

Adiprene B, another elastomer processable on normal rubber machinery, was introduced in 1955. The basic chain is not polyester but polyether in nature; polyethylene oxide has been suggested. This polyether has hydroxyl end-groups and hence its chemistry is very similar to that of the polyesters except that the chain should be less sensitive to hydrolysis. Even so, it is found in practice that uncured compounds have to be stored in closed bins and that for some purposes the maximum storage life of the uncured compounds is only six days. Furthermore, processing has to be kept below 121°C otherwise a granular semi-elastic state is obtained. Carbon black may be used as a reinforcing filler, but the degree of reinforcement which is imparted is very small.

Adiprene C was introduced by the same company two years later to supercede Adiprene B. Processing is even easier than with Adiprene B, and furthermore it resembles natural rubber in being vulcanizable with sulphur and accelerators, or with peroxides. It is suggested that this is due to the presence of unsaturated components in the polyether chain. Pure gum compounds have a fairly low tensile strength, but are stronger than gum vulcanizates

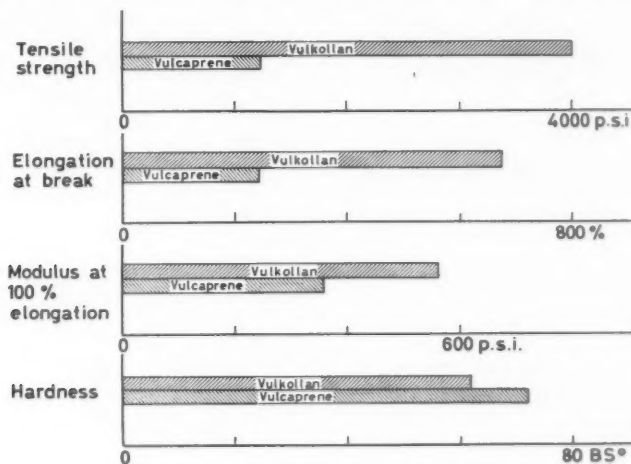


Figure 3. Comparison of Vulkollan and Vulcaprene

from styrene-butadiene rubber. When compounded with 25 parts by weight of carbon black per hundred rubber, a higher tensile strength is obtained than can be obtained with Adiprene B. Discolouration occurs in sunlight, however, so the use of Adiprene C in light coloured compounds is not recommended. Harrington¹ reported that a peroxide-cured Adiprene C compound was more resistant to gamma radiation than the silicone rubbers and fluorine rubbers also examined.

Genthane S is another processable polyurethane, first produced in the United States in 1957. Peroxide curing is suggested, and with dicumyl peroxide carbon black may be used as a reinforcing agent. No special techniques are required during processing, but any fillers used should be free from moisture: several plasticizers are also recommended. Building tack is stated to be adequate for most applications, and adhesion to steel is possible using either isocyanate or rubber hydrochloride bonding agents.

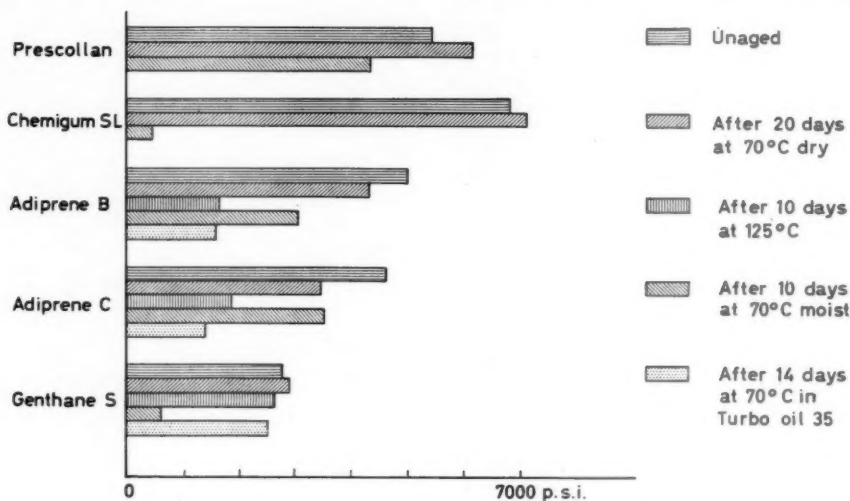


Figure 4. Tensile strengths of polyurethane rubbers after ageing

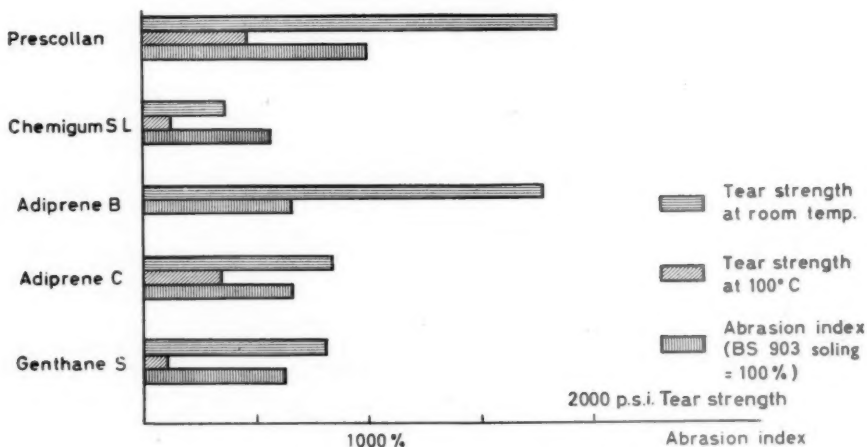


Figure 5. Tear and abrasion resistance of polyurethane rubbers

Properties of Non-Cellular Polyurethanes

In the case of Vulcaprene and Vulkollan the compounds used were:

(i) Vulcaprene A, 100; brown factice, 10; SRF black, 60; stearic acid, 0.75; Vulcafor VHM, 5; Vulcafor VDC, 0.75. Cure: press for 30 minutes at 140°C; oven for 2 hours at 125°C.

(ii) Vulkollan; a transparent sheet supplied by the manufacturers was used for the tests.

The comparison given in Figure 3 shows the superior tensile strength and higher modulus of Vulkollan, although the Vulcaprene is the harder material. This specimen of Vulkollan showed a tear strength of 1650 p.s.i., comparable with that of a natural rubber tyre tread compound whilst a black heel compound showed a Du Pont abrasion resistance ten times that of the natural rubber tyre tread.

The results reported for Prescollan, Chemigum SL, the Adiprenes, and Genthane S have been obtained in these laboratories. Unfortunately the tests have been made at different times over the period 1955 to 1958, so that direct comparison between the elastomers is not possible. Prescollan has been taken as typical of the Vulkollan-type polymers, but it must be remembered that each manufacturer has his own special additives, which impart specific properties to his products. The compounds used were:

(i) and (ii) Prescollan and Chemigum SL; standard compounds prepared by the manufacturer were used. These had hardnesses of 86° and 67° respectively.

(iii) Adiprene B, 100; fine silica, 30; Hylene DMM, 4. Cure: 30 minutes at 134°C and conditioned for 14 days at room temperature and 50 per cent relative humidity.

(iv) Adiprene C, 100; HAF black, 30; coumarone resin, 10; MBTS, 3; MBT, 1; Activator RCD 2098, 0.35; sulphur, 1.5. Cure: 60 minutes at 153°C.

(v) Genthane S, 100; FEF black, 45; stearic acid, 0.2; dicumyl peroxide, 1.6*. Cure: 60 minutes at 160°C.

The results of the physical tests on the above compounds are shown in Figures 4 to 8. Prescollan gives a high tensile strength which is little affected by ageing whether dry or moist (70°C; 100 per cent relative humidity). Chemigum SL is shown to be susceptible to moist heat. The Adiprenes are not very greatly affected by moist heat, but are weakened considerably at 70°C by Turbo Oil 35,

which is a diester lubricant. Genthane is weakened by exposure to moist heat, but resists warm Turbo Oil 35. Neither Prescollan nor Chemigum, which

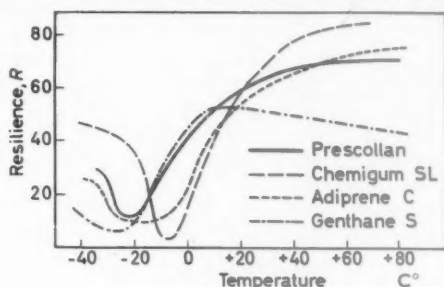


Figure 6. Effect of temperature on the resilience R of polyurethane rubbers

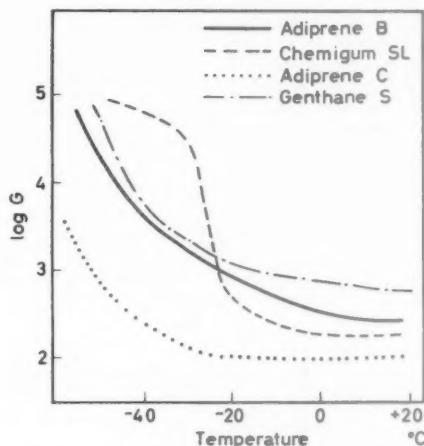


Figure 7. Dependence of the rigidity modulus G of polyurethane rubbers on temperature

were examined earlier in the period to which reference has been made, were tested after ageing at 125°C or after ageing in warm Turbo Oil 35.

Prescollan and Adiprene B show the highest tear strength at room temperature, and Prescollan and Adiprene C the highest at 100°C. The tear strength of Adiprene B at elevated temperatures was not measured. Prescollan shows a very high resistance to abrasion, the remainder have about two-thirds its resistance, but even so they are some three times as good as a natural rubber tyre tread.

Genthane S has a low resilience at high temperatures, but it retains its resilience best at low temperature. Prescollan and Adiprene C show a minimum

* It has been suggested that with 45 parts of black, the dicumyl peroxide content should be increased to 2; this is being investigated in these laboratories.

resilience at about -20°C , but Chemigum SL has its minimum resilience at about -8°C . It is noteworthy that all four polyurethanes examined show roughly the same resilience at room temperature. The rigidity modulus of Chemigum SL shows a sharp increase as the temperature is decreased below -20°C , the others have more gradual rises. This accounts for the high "t10" value of -24°C in the case of Chemigum SL, whereas Adiprene B and Genthane S give -34°C and -41°C respectively, whilst the rigidity modulus of Adiprene C does not attain a value ten times that at room temperature until it has been cooled to -50°C .

Chemigum SL is swollen most by water at 25°C , Prescollan least, whilst Genthane S is the most resistant to mineral oil at 70°C and is quite resistant to isooctane-toluene at 40°C . Adiprene B is swollen most by mineral oil, isooctane-toluene at 40°C , and benzene at 25°C , whilst Prescollan is most resistant to isooctane-toluene and benzene.

It is also noteworthy that all the samples of polyurethane examined were extremely resistant to ozone.

Applications

The above results show that the polyurethanes in general are rubbers of high tensile strength and abrasion resistance, moderately resistant to dry heat, and mineral oil, but susceptible to moist heat; some members of the series show rubber-like properties at quite low temperatures.

Although the high abrasion resistance of polyurethane rubbers at once suggests their use in tyre manufacture, certain drawbacks are found in practice. The heat build-up of these rubbers is said to be so large that it would not be possible to use them in carcase construction and the bonding of a polyurethane tread to a conventional carcase presents many difficulties. Polyurethanes have however been successfully used in solid tyres for relatively slow-moving vehicles (Figure 9): the harder grades may be machined from a rough casting. Here the hardness of the rubber is used to reduce the deflection of the tyre and hence lessen the heat build-up. The oil and tear resistance make it possible to use such tyres in conditions where natural rubber would rapidly become swollen or would be torn to shreds by metal swarf or nails on the floor. In solid tyres, polyurethane rubbers are said to have four times the load-carrying capacity of natural rubber. Motor car components (Figure 10) have also been made of polyurethane rubber.

The properties of polyurethanes mentioned above also suggest their use in belting. The coating of textiles with casting polyurethanes is difficult because of impregnation of the fabric, but there does not seem to be any reason why one of the processable polyurethanes should not be used to make a belt in the normal manner. Polyurethane rubbers have also found use in the electrical industries. Their high abrasion resistances makes the processable kind useful in cable sheathing, whilst the cast polyurethanes are of value as potting compounds.

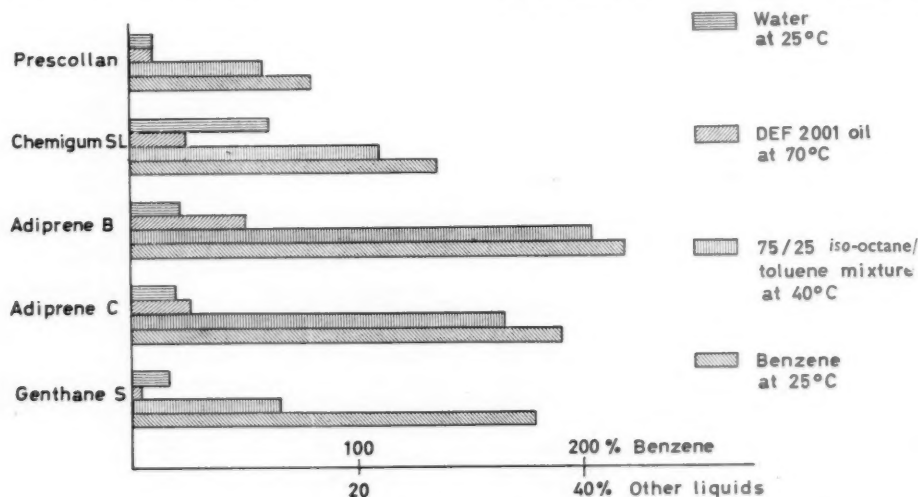


Figure 8. Volume swelling of polyurethane elastomers after immersion for 14 days in various liquids

In footwear, the rather high cost of polyurethane rubbers puts them at a disadvantage, but heels are being manufactured from composite sheets of polyurethane bonded to leather; this construction is necessary for successful nailing to the remainder of the shoe. In the use of polyurethane has been found the answer to the problem of the wear of ladies' 'spike' heels.

The value of polyurethane rubbers in the manufacture of radiation shields has already been noted.

It is in the field of mechanical goods that most use has been made so far of polyurethane rubbers. The extreme abrasion resistance, coupled with moderate oil and heat resistance is of paramount importance here. Printing rollers are one of the important uses of these rubbers, particularly of the softer qualities, and here because of their superior compression set and heat resisting properties they are preferable to the glycerine compositions used hitherto (Figure 11). Another use is in high pressure rubbers, where a hard product of about 95 BS^o is required. There are many applications where part of a machine has to come into contact with abrasive dust or gravel or with chemicals and here it is often possible to use a polyurethane. Pump impellers and diaphragms, the walls of shot-blasting cabinets, the sides of tanks, all may be surfaced with a layer of this versatile material, whilst very hard rigid compounds can be machined to form gears.



Figure 9. Solid tyres made from polyurethane casting polymers (by courtesy of Dunlop Rubber Co. Ltd)

Conclusions

The above account of the properties and applications of solid polyurethane rubbers will suffice to show that they have very valuable properties and are capable of service in a wide range of articles. The highest

physical properties are given by those rubbers which have to be prepared by casting; the processable polymers have hitherto not shown such good tensile properties or such good resistance to tear and abrasion.



Figure 10. Motor car components (by courtesy of Precision Rubbers Ltd)

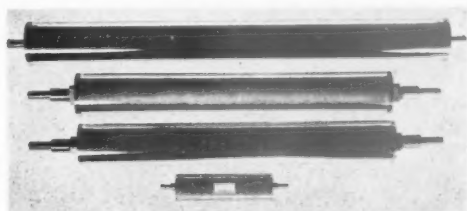


Figure 11. Rollers for printing machinery (by courtesy of I.C.I. Ltd)

There does not seem to be any fundamental reason for this, and doubtless it will not be long before the chemical manufacturers are able to produce a polyurethane rubber which is readily processed on conventional rubber machinery and at the same time has all the properties of the casting rubbers. Such a rubber, whilst it may not command the market of natural rubber or styrene-butadiene rubber, would undoubtedly be welcomed by the rubber manufacturers for a variety of end products, and would form a very valuable addition to the range of polymers now available.

In preparing this article, continual reference has been made to a more extensive review by J. ACRES, published by R.A.B.R.M. Intelligence Division as I.B.C. 442. The laboratory evaluation formed part of the Research Association's general programme of evaluation of raw materials, and the author is indebted to its Council for permission to publish and to colleagues for many helpful discussions.

Reference

- ¹ *Rubber Age* (1957) 82, 462

CARBON BRICKS FOR LINING BLAST FURNACES

D. W. HUBBARD

Carblox Ltd, Loxley, Yorkshire

Carbon bricks are now widely used for lining the well-wall and bosh of a blast furnace. Their use has almost eliminated many previous troubles, namely break-outs, bear formation, and scaffolding. The carbon lining has to be adequately cooled, preferably by spray cooling on the outside of the furnace. Trials are being carried out both in the United Kingdom and in other European countries, using cooling underneath the hearth. Carbon linings can be considerably thinner than the traditional firebrick linings and thereby increase the furnace capacity.

TODAY, carbon bricks are used more or less standard for lining the lower part of a blast furnace, that is the hearth, the well-wall and the bosh. Fifteen years ago carbon bricks were not produced and were almost unthought of. At that time, one firm in this country used a crude form of carbon represented by a mixture of ground coke and tar to patch up the hole in the lining of one of their blast furnaces after a break-out. This mixture was tamped into the hole to stop it but some two years later, when the furnace was stripped, it was found that the mixture had been converted into very hard carbon. This simple action gave the idea to produce carbon bricks to form the actual lining. These were first tried in small patches, then up the well-wall a few feet, then up the bosh. Eventually, four or five furnaces were completely lined with carbon from the extreme bottom to the very top. The latter attempts were only partially successful, but for the lower parts of a blast furnace, carbon has been retained as the best lining refractory.

One of the reasons for the success of carbon in a blast furnace is that its use has just about eliminated the three troubles which, from the beginning of pig iron smelting, have beset the blast furnace. The first and usually the most disastrous of these is 'a breakout'. This occurs when molten iron penetrates the joints between firebricks, usually at hearth level, and travels towards the furnace casing through which it breaks out. This causes a loss of iron and a loss of operating time, both very expensive matters to the iron producer. The second trouble occurs when molten iron penetrates downwards through the joints in the hearth and forms a 'bear' or 'salamander'; this can represent a loss of several hundreds of tons of iron. Another troublesome effect is the disintegration of the ceramic refractory lining of the stack of the furnace near the top leading to scaffolding, chiefly due to the deposition of carbon in the brickwork caused by the dissociation of carbon monoxide in the gases passing up inside the furnace.

Development of Carbon Lining

It has been known for a long time that neither molten iron nor slag will attach to wet carbon. In point of fact the action of molten iron on carbon is similar to that of mercury on wood or glass. Thus it was obvious that theoretically, at any rate, carbon should be the ideal lining for a blast furnace. Before 1939 attempts had been made particularly in Germany to form a monolithic hearth by the use of a coke-tar mixture. This was not entirely successful probably owing to the difficulty of ensuring uniformity during the installation and drying out or curing of the hearth. When such a hearth comes on blast, volatiles will be evolved which cause cracks to develop in the hearth and these may ultimately lead to breakouts or bear formation. In consequence, it was felt that pre-fired carbon pieces would offer better prospects of success as a lining.

Shape of Carbon Bricks

In early efforts to use pre-fired bricks, sizes up to $13\frac{1}{2} \times 6 \times 3$ in. were tried, as these were looked on as a standard blast furnace shape. However, unfortunately there was not sufficient key in this construction to keep the carbon bricks in place and some were found to have floated when the furnace had been in operation some time. Another type tried was the cheese block construction in which blocks were laid in concentric rings around a large centre block, but this was not completely successful.

Eventually a design involving relatively large (approximately $1\frac{1}{4}$ cwt each) blocks, the opposite vertical faces of which were corrugated, was evolved (Figure 1). When properly installed in the hearth of a blast furnace, these blocks are locked together into a firm solid mass. The well-wall and bosh of a blast furnace are best built in the standard blast furnace shapes (see for examples Figures 2 and 3) in carbon, i.e., up to 18 in. long. In certain countries large blocks were initially tried for the

wall but the general tendency is now to use smaller bricks, as more accurate construction can be achieved by their use.



Figure 1. Corrugated block hearth in a blast furnace

Depth of Carbon Lining

In general, the depth of carbon in a blast furnace hearth should be 75 in. The well-wall may be from 22½ in. to 36 in. depending on the hearth diameter. The bosh is generally 18 in. to 22½ in. It must be pointed out here that these thicknesses for well-wall and bosh can be smaller than is the case if the lining be constructed with ceramic bricks. In earlier days even a small furnace would have a well-wall from three to four feet thick. Now the same sized furnace would have a carbon well-wall only 22½ in. or 24 in. wide. These same remarks also apply to the thickness of the bosh lining. Thus by the use of carbon, the effective hearth diameter can be increased and in consequence a greater quantity of iron can be produced from the same size of shell. This greatly assists productivity—so essential these days. Figure 3 shows a carbox bosh whose thickness has remained about the same after four years operation.

Around the carbon lining, whether in brick or block form, it is advisable to install approximately 1½ in. to 2 in. of carbon ramming material between

the outside of the carbon brickwork and the inside of the casing. This ramming fills up any irregular spaces, also acts as a cushioning by taking up any movement in the brickwork due to expansion etc. It also ensures good contact between the carbon lining and the cooled shell, and thus utilizes the high conductivity of carbon to the maximum. (Carbon is also used for the runner trough; see Figure 4.)

Cooling of the Blast Furnace

The thermal conductivity at any temperature of carbon is about three times as high as that of fire-brick. Thus it is essential that the blast furnace with a carbon lining be adequately cooled, as only then can advantage be taken of the special properties of carbon. In most modern furnaces the older fashioned types of cigar and plate coolers are being replaced by external spray cooling. This is much easier to install and much more economical, as it avoids the use of very expensive copper coolers. It



Figure 2. Hearth and wall of No. 6 furnace, blown in April, 1949—blast furnace lined with carbon standard shapes

has the additional advantage that the water is in its correct place, i.e., outside the furnace, cooling the casing. With cigar or plate coolers, if one becomes burnt or broken, then there will be a leakage of water into the lining. Although there will be a relatively steep temperature gradient

through the lining, certain parts of the latter will be at the critical temperature at which steam or water vapour will oxidize carbon. Throughout, it must be remembered that carbon can only be successful if used in a reducing atmosphere. It must not be used in an oxidizing atmosphere.



Figure 3. Carblox bosh which is still the same thickness after four years' operation



Figure 4. Carblox runner trough at Renishaw

External spray cooling is largely replacing the use of coolers in the well-wall and bosh and stave coolers around the hearth. To ensure the adequate cooling of the hearth proper, *i.e.*, the centre pad thereof, various forms of cooling have been tried and adopted. In Sweden, water and oil (separately) have been circulated in coils below the hearths of furnaces, air has similarly been used in Finland, also in Sweden. In Austria, at one plant, there are two blast furnaces literally standing in water, and one of these has operated like this for many years without a break-out. In this country, a furnace at one well known works has its hearth cooled by blowing air in opposite directions through channels under the carbon hearth lining. This furnace is working very efficiently and the temperature at the bottom of the carbon never varies far from 240°C.

Height of Carbon Lining

It is undeniable that properly cooled carbon is the best refractory for lining the lower half of a blast furnace. The same cannot yet be said for the lining of the stack of a blast furnace. As mentioned above, certain furnaces have been tried with complete carbon linings. Whilst these eliminated some troubles, they were not completely successful. At the present time certain special ceramic refractories are being tried in stacks. These have yet to be proved in operation. In certain cases where exceptional troubles have been experienced, carbon has proved beneficial in the lower half or third of the stack. Alkali attack is an example of this.

The advantage of using carbon to line a blast furnace from the top of the bosh downwards (see, for example, Figure 5) is that it eliminates or greatly reduces the troubles mentioned above which always have occurred in blast furnaces. At one of the biggest iron producing works in this country, probably owing to having to work with lean ores, there were, before carbon was used, the largest number of breakouts of any works in this country. Since carbon linings have been used there has not been a breakout through carbon. This is true of nearly all other ironworks using carbon linings in their blast furnaces. Scaffolding was eliminated when carbon was used to line the stack and the same appears to be true now that special high temperature fired ceramic bricks are being mostly used in stacks. The use of carbon for the centre pad of the hearth of a blast furnace had not completely eliminated the formation of bear but it has greatly reduced the amount of iron which can be tapped from below a furnace when being blown out. Every operator of blast furnaces knows that when using carbon linings there will be very many less stoppages than when as previously, firebrick is used.



Figure 5. Appelby-Frodingham No. 11 furnace (one of the largest in Britain) which is lined with Carblox up to the top of the bosh

Other Applications of Carbon

Carbon is frequently used in case of trouble or difficulty. A typical example of this is found in the steel industry where in recent years pickling or descaling of stainless steel has been carried out by using a mixture of nitric and hydrofluoric acids. No ceramic or other acid resisting materials will withstand this mixture for long but carbon can be used. In consequence carbon bricks are now used exclusively for lining such pickling tanks except in the case of electrolytic pickling. (For electrolytic pickling the anodes can cause disintegration of parts of the carbon lining.)

Whilst referring to the acid resisting properties of carbon it may be mentioned that carbon is very extensively used for lining sulphite paper pulp digesters. The latter are the large mild steel vessels in which wood chips are heated under pressure in the presence of sulphurous acid, or more commonly, calcium bisulphite—whence the name of this particular process as compared with the other two usual processes, namely, mechanical or sulphate. For many years in the sulphite mills, digesters were lined with one or two layers of ceramic acid resisting tiles. But after these had been in continuous opera-

tion for two years or more, it was found that the inner faces of the ceramic tiles began to flake off. In consequence, carbon tiles are now used throughout Europe—including Norway, Sweden and Finland, the largest producers in the world of paper pulp, also Austria. Carbon is also extensively used for lining fertilizer plants because, unlike ceramic material, it will withstand phosphoric acid.

The use of carbon is by no means confined to this country. Its use has been developed in America quite extensively since 1944. It has mostly been installed there in the well-wall and hearth but it is now beginning to be used also in the bosh. Carbon has been used in Germany for over sixty years but only recently has the use of pre-fired bricks or blocks become standard practice. One type of carbon made in England has been and/or is being used in blast furnaces in Sweden, Finland, Belgium, France, Spain, Austria, Yugoslavia, Hungary, Australia, South Africa, Rhodesia and Canada. This particular carbon has the highest cold crushing strength and highest abrasion resistance of any carbon made anywhere in the world.

Carbon has undoubtedly made a marked contribution to the ever increasing production of pig iron

THE MOLECULAR BASIS OF LIFE

M. F. PERUTZ

Medical Research Council Unit for Molecular Biology, Cavendish Laboratory, Cambridge

This article is based on a Discourse delivered at the Royal Institution on May 20th, 1959, and it discusses how the laws of inheritance are now beginning to be understood on a molecular basis.

LAST year we celebrated the centenary of Darwin and Wallace's reading of a paper in which the concept of evolution was first formulated. The idea that the tremendous variety of different living forms might have developed by gradual modification and natural selection from a common primeval ancestor seemed preposterous as well as heretical to most of Darwin's contemporaries. How could a flower and a man have both developed from the same amoeba? The idea of a common origin looks less absurd if the flower and the man are both cut into thin slices and examined under the microscope. The tissues of both organisms are divided into cells. When these are about to divide they are seen to contain circular regions enclosed by a thin membrane. Inside this membrane are bodies with a variety of shapes which can be stained with basic dyes and are called chromosomes or coloured bodies.

The research of a generation of scientists has proved the chromosomes to be the carriers of the cell's inheritance, particular inherited characteristics, like colour of the eyes, being associated with particular chromosomes. By studying the results of crosses between genetically different individuals, research workers actually traced the location of particular characters to minute regions arranged along the chromosomes in a linear order, like the beads in a rosary. These regions, which are associated with specific inherited traits, are known as genes.

The Role of Nucleic Acid

The similarity which strikes us when we look at widely different organisms on the microscopic scale becomes even more apparent when we examine the actual molecules responsible for their propagation and growth. Chemical analysis shows that all chromosomes, from whatever organism, consist of the same two chemical compounds: nucleic acid and protein; clearly, therefore, one or the other of these compounds must be the carrier of inheritance.

The first question to be answered is which of the two is really essential. Experiments on micro-

organisms have now furnished proof that it is the nucleic acid alone which carries the genetic information. The most convincing answer comes from A. D. HERSHEY and M. CHASE's experiment on the replication of a virus which is a parasite of coli bacteria and is called T2 bacteriophage. This consists of a head containing a great length of nucleic acid enclosed in a protein shell, and a double-sheathed hollow tail (*Figure 1*). The virus attacks a bacterium tail first, punctures the cell wall and within a few seconds injects the whole of its nucleic acid into the bacterium, while the protein shell remains outside, dead like the bee after its sting. The injected nucleic acid at once reorganizes the bacterial metabolism to the single purpose of producing bacteriophage. Within a few minutes after infection a great pool of phage nucleic acid is synthesized; this crystallizes to form hexagonal heads which surround themselves with a protein shell, and grow tails. Within half an hour the bacterium bursts with the release of several hundred progeny phage particles, ready for their next victims. The fact that only the nucleic acid enters the bacterium, while all the protein remains outside, proves that the nucleic acid alone carries the heredity of the virus.

The nucleic acid of phage is of a kind called deoxyribonucleic acid or DNA, and is the carrier of heredity in the larger viruses and in all higher organisms. Certain small viruses contain a nucleic acid of a slightly different kind called ribonucleic acid or RNA.

An example is the mosaic virus which causes mottling in the leaves of tobacco plants. X-ray analysis by A. KLUG and the late ROSALIND FRANKLIN has shown that it consists of a single helical chain of RNA surrounded by a symmetrical array of protein molecules, all exactly alike (see *Figure 2*). A. GIERER and G. SCHRAMM in Germany and H. FRAENKEL-CONRAT in America stripped the protein off the RNA and rubbed each of them separately into the leaves of different tobacco plants. They found that only the RNA was infective, but not the protein.



Figure 1. Electron micrograph of T2 bacteriophage

There is no such direct proof as yet of the part played by nucleic acid in the chromosomes of higher organisms. However, the constancy of the DNA content in all cells of a given species, its metabolic inertness and a variety of other evidence make us believe that DNA is in fact the sole carrier of heredity in all organisms except the small viruses where its place is taken by RNA.

The Genetic Code

If the nucleic acids are to convey genetic information they must be so made as to form a chemical code. They are in fact long-chain polymers in which an identical chemical pattern repeats at regular intervals, forming the links of the chain. Attached to each

link is a purine or pyrimidine base. In any one nucleic acid these bases are of four different kinds (Figure 3). We do not yet know in what sequence the bases are arranged in any nucleic acid, nor have we any direct information that they are arranged in any definite sequence at all; we know only that their proportions are characteristic and constant in the DNA of any particular species. It is the fact that the bases are the only variable constituents that makes us believe that they are arranged in a definite sequence and that this sequence constitutes the genetic code. If this is true then the genetic language is written in a four letter alphabet on an immensely long scroll. The actual number of letters in the RNA chain of tobacco mosaic virus

is 6000, in the DNA of T2 bacteriophage 500,000 and in the chromosomes of a mammal about 3000 million.

To ensure continuity of inheritance, an exact copy of this information has to be made each time a cell divides. In 1953, J. D. WATSON and F. H. C. CRICK

proposed a structure for DNA which suggests a possible copying mechanism (Figure 4). It consists of two chains of DNA coiled round each other, like two snakes, to form a double helix. The actual model looks like a spiral staircase in which the links of the nucleic acid chain form the banisters, and the bases attached to them form the steps. Each step consists of two bases, one from each of the chains, which are linked together by chemical bonds. Suppose now that the four bases which form the symbols of the genetic code are called A, T, G and C, then the model predicts that only specific pairs of bases can be linked to form a step, such that A is always linked to T, and G to C. This means that a particular sequence of bases in one chain must be paired to a complementary sequence in its partner chain (Figure 5). This is the vital idea of Watson and Crick's model, for from it we can deduce the way in which the genetic code might copy itself.

Method of Reproduction

When we want to copy a document we prepare a negative from which we make a positive print. The complementary sequence of bases in the two chains make DNA a duplex of negative and positive combined. To reproduce themselves the two chains of the parent double helix would have to separate in a solution full of loose chain links, and each chain would have to become a template for the formation of a new complementary chain growing around it. This means that a loose chain link carrying adenine joins on to thymine in the parent chain, a loose chain link carrying guanine joins on to cytosine in the parent chain, and so on. When all the loose chain links are joined, each parent chain will be intertwined with a new daughter chain made up of a complementary sequence of bases. In this way one parent double helix gives birth to two daughter double helices each carrying an exact copy of the genetic information (Figure 6).

The helical structure of DNA, the idea of the genetic code as a definite sequence of four different kinds of chemical groups along a pair of intertwined

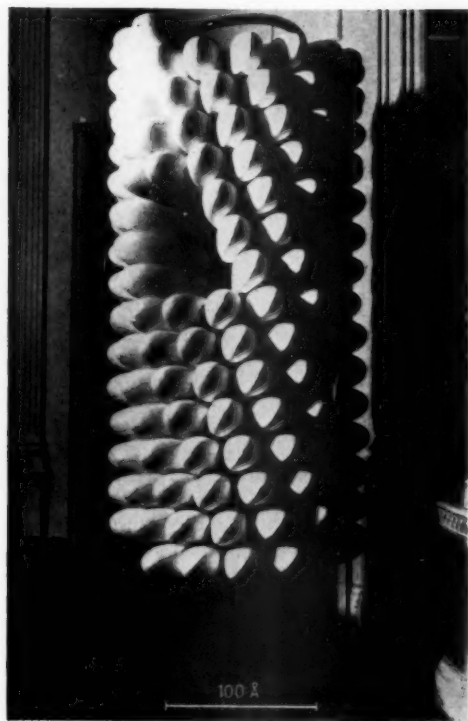


Figure 2. Model of a short length of the tobacco mosaic virus particle built by A. KLUG and the late ROSALIND FRANKLIN. The white leaves represent protein molecules. In one place these have been cut away to show the helical rod representing the RNA chain underneath.

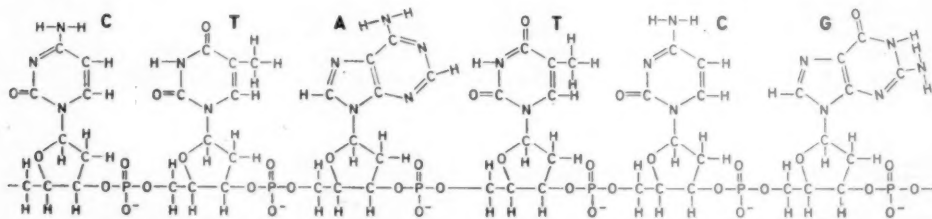


Figure 3. The chemical constitution of DNA. The deoxyribose-phosphate ester chain is at the bottom. At the top are the nitrogenous bases: C=cytosine, T=thymine, A=adenine, G=guanine

chains, and the pairing of the groups with its implied copying mechanism, belong to the most beautiful scientific ideas of our time. What is beautiful is not necessarily true, however, and scientists, trained to be sceptical, awaited experimental verification. M. H. F. WILKINS and the late Rosalind Franklin soon proved by x-ray analysis that the helical model of DNA was right, but the

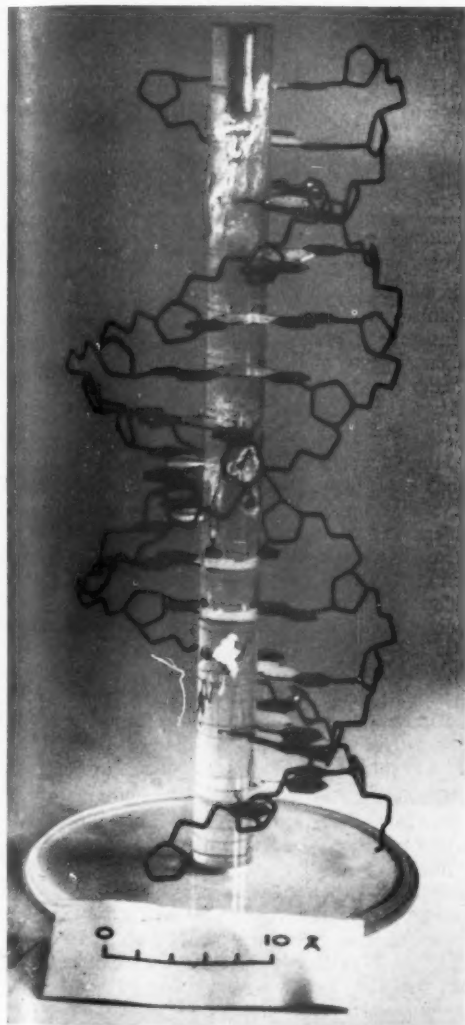


Figure 4. Model of the double helix structure of DNA. The phosphate ester chain is represented by wires and the bases by discs

idea of a self-replicating molecule was so new that it took scientists several years to test it.

One ingenious test consists in allowing micro-organisms whose DNA has been isotopically



Figure 5. Base pairing in DNA

labelled to reproduce in the presence of normal, unlabelled precursors. If the Watson-Crick mechanism is valid then the parent double helices would uncoil, and each of their labelled chains would serve as a template for the formation of a new chain having no isotopic label (Figure 7). The offspring of the first generation would thus contain mixed double helices, each carrying a labelled and an unlabelled chain. During the growth of the second generation these would segregate into two types: one mixed and the other unlabelled. In further generations the unlabelled double helices would increasingly predominate over the mixed ones. Several ingenious experiments on these lines have now been done and each has given an answer consistent with the mechanism of replication indicated in Figures 6 and 7.

The Role of DNA

Having discussed the structure and replication of the genetic material we now turn to its purpose. What does it do? If we compare a living organism to the body-politic, the genetic message can be regarded as its constitution, laying down the form which the state must take. What is the machinery of government which translates it into action? We now know that each gene controls the synthesis of one specific enzyme. Each enzyme, in turn, catalyses one specific step in one of the many thousands of chemical reactions involved in the life of an organism. Indirectly, therefore, each small chemical step is controlled by a gene.

While genes are made of nucleic acids, all enzymes discovered to date are proteins. It must be concluded that it is the purpose of a gene to control the structure of a protein molecule.

Like the nucleic acids, proteins are giant molecules. They consist of polypeptide chains in which the same chemical groups

-CO-RCH-NH-CO-RCH-NH-

repeat at regular intervals. Specificity is provided by

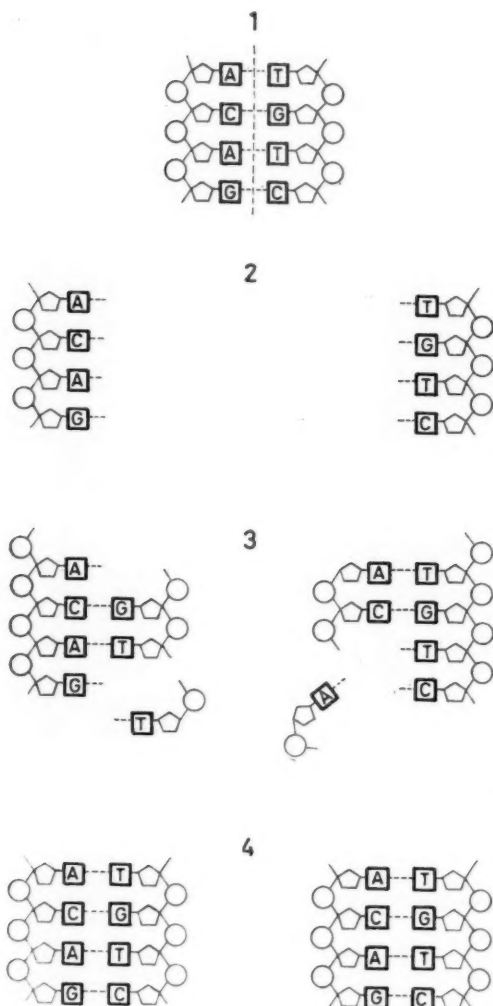


Figure 6. Possible replicating mechanism of DNA. The two linked chains (1) come apart (2). Monomers then assemble along each chain (3). The result is two pairs of linked chains where bases (black squares) have the same sequence as those of the original pair. (Reproduced by permission from F. H. C. CRICK, *Scientific American* (1957), page 88)

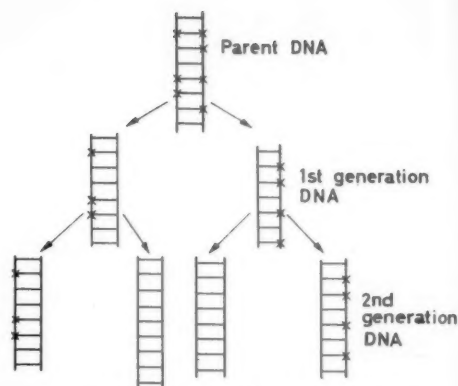


Figure 7. Distribution of isotopic label of parent DNA among double helices of progeny DNA. The crosses indicate isotopic markers. (Reproduced by permission from Endeavour (1958), 17, 90)

the side chains *R* of the twenty different kinds of amino acids (Figure 8). Insulin, for instance, consists of 51 amino acid residues in two polypeptide chains. Myoglobin consists of a single polypeptide chain contorted in a strange fashion to form a kind of basket around an iron-containing pigment group (Figure 9).

How does DNA determine protein structure? The unique feature which proteins and nucleic acids have in common is a specific linear sequence of different units on a long chain. What distinguishes them is the number and kind of units which each contains. Any one DNA contains mainly four different kinds of bases, whereas proteins contain mainly twenty different kinds of amino acids, and it is the sequence of these which determines their catalytic activity. We now assume as a working hypothesis that the sequence of the four bases along the nucleic acid chain determines the sequence of amino acids along the protein chain, and that this is the real purpose of the genetic code.

How can a four symbol code be made to express words in a twenty letter alphabet? It can be shown that a sequence of at least three DNA letters would be needed to code for one amino acid. One difficulty is that the genetic message is not allowed to have any gaps or commas. How do we know where one triplet ends and the next one begins? F. H. C. CRICK, J. S. GRIFFITH and L. E. ORGEL have solved this problem by devising a code in which only twenty triplets make sense and all others make nonsense (Figure 10).

We do not yet know whether the genetic code is of this kind. One of the weaknesses of the triplet code

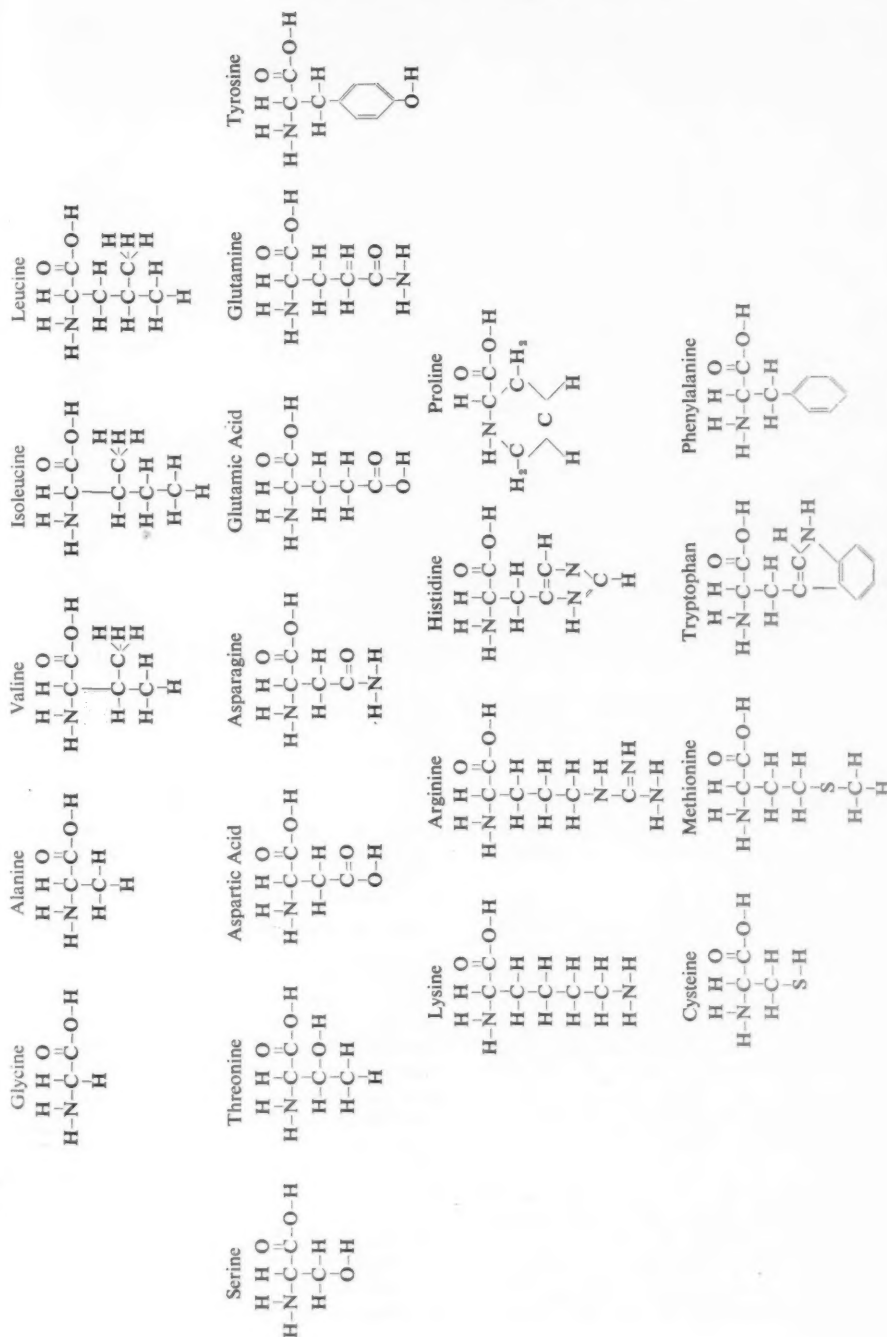


Figure 8. The twenty commonly occurring amino acids. Note that the eight atoms in the top two lines from which the main chain is formed are common to all amino acids, while the other atoms differ in each. These different chemical groups are known as the 'side-chains' or R groups



Figure 9. J. C. KENDREW's model of the myoglobin molecule. The light winding rod represents the polypeptide chain, the dark disc the haem group

is that it can be read in either direction, so that the polypeptide chain would not know whether it is to be assembled forwards in the direction of the carboxyl end or backwards in the direction of the amino end. To give it direction, a quadruplet code would be required.

Whatever the nature of the code, we must have a chemical translation mechanism. There must be devices for holding the template rigid, for bringing the amino acids to the assembly line and hooking them together in the right order. Great strides have been made in the last few years in elucidating the sequence of chemical reactions involved in this process. Whatever the precise nature of the DNA code, it is now widely believed that proteins are made on nucleic acid templates where the amino acids are assembled in the right order before being linked together to form polypeptide chains.

From what has been said so far one would naturally expect the templates for protein synthesis to occur in the cell nucleus. In fact, most protein is made outside the nucleus, where the point of assembly has been tracked down to small particles consisting of about half RNA and half protein.

We have reason to believe that it is the RNA in these particles which form the templates on which protein chains are made. This implies that the coded message for the assembly of the protein molecule is transferred from the DNA in the nucleus to RNA and that this RNA migrates into the cytoplasm. About this transfer nothing is as yet known.

When the idea of nucleic acid as a template for protein synthesis was first discussed a few years ago it appeared stereochemically improbable, because there is no way of arranging purine and pyrimidine bases in twenty different ways so that each combination of bases has a specific affinity for one particular amino acid. In other words, it was not possible to discover any direct lock and key relationship between proteins and nucleic acids, such as would be expected if proteins were to be made on nucleic acid templates. Yet all the experimental evidence pointed in that direction.

To get over this dilemma F. H. C. CRICK made the ingenious proposal that each amino acid might be brought to the template by a specific carrier. He suggested that this carrier might be a short length of RNA chain with a specific sequence of bases coding

for the particular amino acid. The carrier would bring the amino acid to the right point on the assembly line by joining it on to a complementary sequence on the RNA template. All that would be needed to make this mechanism work are specific enzymes which join each kind of amino acid to the particular carrier which contains its code word.

Recent experiments have confirmed this hypothesis. Each amino acid is activated by a specific enzyme which first joins it to adenosine monophosphate (AMP) and then links the resulting amino acid-AMP-anhydride

on to a relatively short, soluble, piece of RNA chain which may be the carrier postulated by Crick. As was to be expected, there appear to be twenty different activating enzymes, each specific for one particular amino acid, but we do not know yet whether there also are twenty different pieces of soluble RNA chain, each acting as a carrier for one particular amino acid. Experiments to test this point are under way.

The mechanism of protein synthesis is not yet understood, but it is being explored by biochemists all over the world, and new details may well become known before this article appears.

Variations in Inheritance

From what we have learnt so far, the living world would remain static. The interchange of genetic material which takes place in sexual reproduction would forever produce new combinations of those genes which already exist, but as long as the DNA templates are copied faithfully from generation to generation no new genes could arise. How could evolution be explained in such a system? How is this to be reconciled with Darwin's teaching that new varieties of plants and animals arise spontaneously?

Experiments on microorganisms have shown that new varieties do arise and are usually distinguished from the normal or wild type by a change in one specific enzyme. An existing enzymatic activity may be suppressed or modified, or a new activity may be added. For instance a culture of streptococci may give birth to a new and deadlier variety which is penicillin-resistant because it has an enzyme capable of destroying penicillin. Results of this kind have led us to believe that a genetic change, or mutation as it is generally called, expresses itself in the first instance by a change in the structure of a protein molecule.

This expectation has been verified in a beautiful series of experiments by V. M. INGRAM and J. A. HUNT. Haematologists have recently discovered a variety of congenital anaemias which are inherited

in a Mendelian manner. Each of these diseases is now known to be due to the production of a wrong protein molecule. Haemoglobin, the protein in the red cell which transports oxygen from the lungs to the tissues, is replaced by a protein which is somehow abnormal and alters the properties of the red cell in a deleterious manner. Ingram and Hunt have discovered several of these abnormalities to be due to the change of a single one out of the 300 amino acid residues in the half-molecule of haemoglobin (Figure 11).

Looked at from the point of view of protein chemistry, these discoveries show us that the

The four symbol triplet code

A	T	A	A	A	A	T	G	C	G
		T	C	T	T	C			

There are 20 different triplets. Each can be assigned a small letter

ATA	ATT	ACA	ACT
a	b	c	d

and so on

These five triplets make sense

s	e	n	s	e
T	G	C	A	C
C	C	T	G	A
T	G	C	A	C
?	?	?	?	?

These triplets make nonsense

Figure 10. The triplet code of Crick, Griffith and Orgel. Each letter of a twenty alphabet can be symbolized by a sequence of three of the four letters A, T, G and C. If it is assumed that only the twenty combinations indicated at the top of the figure make sense, then the code can be commaless, in the sense that no intermediate sequence would make sense. This is indicated at the bottom of the figure

alteration of a single residue sensitively placed in a protein molecule may have a profound effect on its physiological properties. Genetically, they teach us that mutations work on an extremely fine scale, one mutation changing only one residue in a large protein molecule.

What is the origin of these changes? If our picture of the structure and function of the genetic code is right, then the change in a single amino acid residue in the protein should be due to the replacement of a single base in the DNA code. Such a mistake could occur if a wrong base were incorporated in one of the daughter chains during

HbA . . . Val-His-Leu-Thr-Pro-Glu-Glu-Lys . . .
 HbS . . . Val-His-Leu-Thr-Pro-Val-Glu-Lys . . .
 HbC . . . Val-His-Leu-Thr-Pro-Lys-Glu-Lys . . .
 HbG . . . Val-His-Leu-Thr-Pro-Glu-Gly-Lys . . .

Figure 11. Sequence of amino acid residues in a small segment of one of the polypeptide chains of haemoglobin. The three rows show the sequence in normal human haemoglobin (HbA), in sickle cell haemoglobin (HbS), in Haemoglobin C disease (HbC), and in haemoglobin G disease (HbG); the last has been elucidated by R. L. Hill and H. C. Schwartz. It is seen that in each of the diseases all the residues except one (in *italics*) remain unaltered

replication. This could alter the code so that it spells 'valine' at the point where it spelt 'glutamic acid' before, with the result that the chemical properties of the haemoglobin molecule synthesized by this gene are adversely altered.

There is no direct proof as yet that this is how it happens, because the human genetic apparatus is too complex to follow the replacement of a single DNA base in one of its chromosomes. In tobacco mosaic virus, on the other hand, whose gene consists of a single RNA chain of 6000 residues, A. GIERER has been able to show that the oxidation of a single amino group in any one of 180 out of the 6000 bases gives rise to a mutant virus producing yellow lesions in a plant in which normal virus produces green lesions.

Summary

The laws of inheritance which have been studied empirically since Darwin and Mendel's time are now beginning to be understood on a molecular basis. The character of a species is laid down in a linear

code which consists of the sequence of four symbols along a nucleic acid chain. This code reproduces itself exactly each time a cell divides. It serves the purpose of determining the sequence of the twenty different kinds of amino acids in the polypeptide chains of proteins. The proteins in turn are the enzymes which catalyze the chemical reactions on which life depends. Thus, indirectly, the genetic code controls the kinds of chemical compounds from which the organism is to be built and the reactions which give it the energy to live.

We suspect now that this is not the whole story, because there must also be a mechanism controlling the sequence in which the different enzymes are brought into play, since otherwise it is hard to see how differentiation is accomplished on which the growth of higher organisms depends. This is perhaps the most important field for the future of molecular biology, and it is in that direction that the solution of the cancer problem may have to be sought.

The nucleic acid code replicates itself by a template mechanism each time a cell divides. If a mistake occurs during replication and a wrong base is incorporated in one of the daughter chains, then the genetic message is *permanently* altered. This might make the message meaningless, thereby inhibiting the synthesis of a vital enzyme. Such a mutation would be lethal, in which case no further reproduction would occur; but if the mutant organism is viable then the template mechanism of DNA replication would cause the altered genetic message to be transmitted indefinitely from generation to generation, thereby giving rise to a new variety in the species. Thus evolution would proceed in a series of small steps, each involving the change of a single quantum of heredity, that is to say of one single base. By altering the sequence of amino acid residues in a protein such changes would lead to the evolution of new kinds of enzymes. These in turn would allow the organism to catalyze new kinds of chemical reactions and eventually lead to the development of new species.

See also item on the Cell Nucleus Symposium on page 361

THE PIGMENTS OF BLACK APHIDS

R. I. T. CROMARTIE

University Chemical Laboratory, Cambridge



Figure 1. Colonies of *Tuberolachnus salignu* on the bark of a willow tree

MANY species of aphid, especially the dark-coloured ones known to farmers and gardeners as 'black fly', contain a red pigment that undergoes a remarkable series of changes when the insects are crushed and exposed to the air. This phenomenon was first observed on woolly aphids as long ago as 1871 by H. C. SORBY¹, who showed that the original water-soluble pigment was transformed successively into three ether-soluble, fluorescent substances: they were, in order of appearance, yellow, orange and red; and each was distinguished by characteristic sharp absorption bands in the visible spectrum. In 1936 B. K. BLOUNT² isolated two of these derived pigments from different species of insects and described their properties: from the woolly aphid *Eriosoma lanigerum* he obtained the orange lanigerin and from the pine chermes *Adelges strobi* the red strobilin, both of which he recognized as hydroxyquinones. (Examples of the species of aphid, *Tuberolachnus salignu* on the bark of a willow tree are shown in Figure 1.)

The Series of Aphins

A thorough chemical investigation of this series of pigments was begun by A. R. TODD and his collaborators in Cambridge in 1947 and has elucidated some of the surprisingly complex chemical problems connected with them. The living insect contains a halochromic pigment protoaphin, which has a deep crimson colour in neutral or alkaline solution, changing reversibly to yellow below pH 5.5. It can only be isolated from aphids that have been killed by scalding with water at 70° to destroy an enzyme present in the insect which converts it rapidly in the presence of air into the three ether-soluble pigments observed by Sorby. The initial product of the enzymic reaction is the yellow xanthoaphin, which is very unstable, passing under the influence of heat or light or mild acidic or basic treatment into the

orange chrysoaphin. Chrysoaphin is more stable, but rather more vigorous acidic or basic treatment transforms it into the stable end-product of the series, the deep red erythroaphin. Extraction of fermented aphids usually yields a mixture of all three, which can be separated into its components by fractional crystallization^{3,4}.

Occurrence

Pigments of these types, collectively known as aphins, have been detected in more than twenty different species of aphid, belonging with two exceptions to the family Aphididae, though they are present in very different amounts⁵. The richest source so far examined is the willow aphid *Tuberolachnus salignus*, which contains up to two per cent of its live weight of pigment; it is found in dense colonies on the bark of willow trees in the autumn and can be collected alive in large numbers.

The second main source used for extraction of aphins has been *Aphis fabae*, the well-known black aphid that frequently infests broad beans in the summer; it contains much less pigment than the willow aphid, but it can be collected by the kilogram in a favourable season. It is fortunately unnecessary to work up the aphids at once, unless xanthoaphin is required, since they can be preserved for years below -10°C without much loss of pigment; but they must first be killed with hot water if protoaphin is to be extracted later.

The aphins from different aphids are designated by suffixes indicating the species e.g. erythroaphin fb from *Aphis fabae* and erythroaphin sl from *Tuberolachnus salignus*. On this nomenclature Blount's strobilin is erythroaphin st and his lanigerin is chrysoaphin ln. The corresponding aphins from different species are closely similar but at least two distinct series of isomeric compounds can be

recognized, the aphins fb and the aphins sl. It is probable that all the other natural erythroaphins are identical with either erythroaphin fb or erythroaphin sl.

dihydroxyperylene-3:10-quinone (Figure 2d) by synthesis of this substance, which was already known but had not been recognized as such: it and the erythroaphins had closely similar spectra as did the

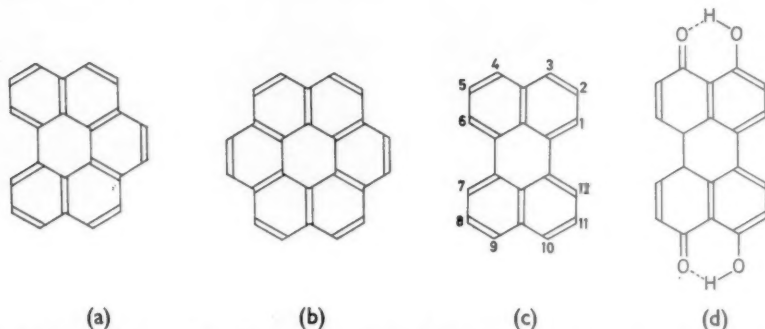


Figure 2. (a) 1:12-benzperylene; (b) coronene; (c) perylene and (d) 4:9 dihydroxyperylene-3:10-quinone

The Chromophore of Erythroaphin

In view of the instability of the protoaphins, xanthoaphins, and chrysoaphins the erythroaphins have been studied most extensively: it was considered that a firmly established constitution of the stable end-product would afford a firm foundation on which to build subsequent deductions of the nature of its precursors.

Erythroaphins fb and sl are isomers of molecular formula $C_{30}H_{12}O_4$ with almost identical visible and ultra-violet spectra; they differ in their infra-red spectra and in some other physical properties. On hydrogenation they take up two atoms of hydrogen to give a yellow vat, readily reoxidized to erythroaphin by air. Acetylation yields a diacetyl derivative and reductive acetylation a tetraacetyldihydro derivative, showing that erythroaphin is a dihydroxy-quinone^{3,4}. Evidence from infra-red spectra suggested that erythroaphin is an 'extended quinone', i.e., that the carbonyl groups are in different rings; and that the hydroxyl groups are strongly hydrogen bonded to them⁵. Most of the usual degradative processes yielded either no well defined product or small fragments of little value for structural studies; but fusion with zinc dust, in a salt melt, yielded a mixture of polycyclic aromatic hydrocarbons. From the combined products of 250 such experiments three different ring systems were separated by chromatography and identified spectroscopically: alkyl derivatives of 1:12-benzperylene (Figure 2a) and coronene (Figure 2b) were isolated as crystalline picrates; and homologues of perylene (Figure 2c) were also present⁷. The common chromophore of the erythroaphins was identified as that of 4:9-

products of their reductive acetylation, though the aphins spectra were shifted bathochromically relative to the model compounds⁸.

The Substitution Reactions of Erythroaphin

The erythroaphins undergo an interesting series of substitution reactions. They react with ammonia in presence of air by alternate addition and reoxidation to give diamino derivatives. Bromination leads to dibromo derivatives, which are still capable of reacting with ammonia to give dibromodiamino derivatives. By treatment with nitrous acid the amino groups can be replaced by hydroxyl groups, which can also be introduced into erythroaphin by direct oxidation. All these substituents can be removed by reduction with zinc dust and acetic acid.

Whereas erythroaphins fb and sl give distinct dibromoderivatives, the derivatives containing hydroxy and amino groups prepared from the two erythroaphins under basic conditions are identical; and reduction of such compounds always yields erythroaphin fb⁹. Thus erythroaphin sl can be converted into the fb isomer in two steps, which confirms the very close relation between the two isomers; the same transformation can be effected directly by treatment with sodium hydroxide, though in very poor yield.

The Constitution of Erythroaphin

The properties of dibromodiaminoerythroaphin suggest strongly that the four substituents are all attached to the quinonoid nucleus, leaving four vacant positions for the attachment of the residue $C_{16}H_{14}O_4$ that is as yet unaccounted for. The

inertness of these four oxygen atoms suggests that they are present as cyclic ethers, and since this residue is saturated it must form four additional rings. The formation of coronene derivatives on degradation with zinc dust shows that of the 'waist' positions 1, 6, 7 and 12 of the perylene nucleus at least one on each side must bear alkyl substituents.

Two small but significant fragments have been obtained by degradation of the aliphatic part of the molecule. Heating with 60 per cent sulphuric acid liberates about 0.8 mole of acetaldehyde from erythroaphin sl and almost exactly twice as much from erythroaphin fb. On the other hand oxidation by the Kuhn-Roth procedure, which degrades methyl groups attached to carbon to acetic acid, produces the same yield of 3.8 moles from both isomers. These facts can be explained in terms of four alternative formulae for erythroaphin (Figure 3). The isomerism of the erythroaphins can be explained on

The reverse change can be effected photochemically: if the product of reductive acetylation of either erythroaphin is irradiated with ultra-violet light the same mixture of the two compounds is formed, which can be hydrolyzed and oxidized to a mixture of erythroaphins. This mixture also contains small amounts of a third isomeric erythroaphin, which has never been obtained from any species of aphid; erythroaphin sl is intermediate in properties such as solubility and optical rotation between erythroaphin fb and the third isomer, which suggests that the latter has *trans*-fusion on both sides, thus completing the series of possible isomers of this type¹¹.

The choice between the two alternative formulae is very hard to make on the basis of chemical evidence but recently the novel method of nuclear magnetic resonance spectroscopy (N.M.R.) has provided strong evidence in favour of the dioxolan

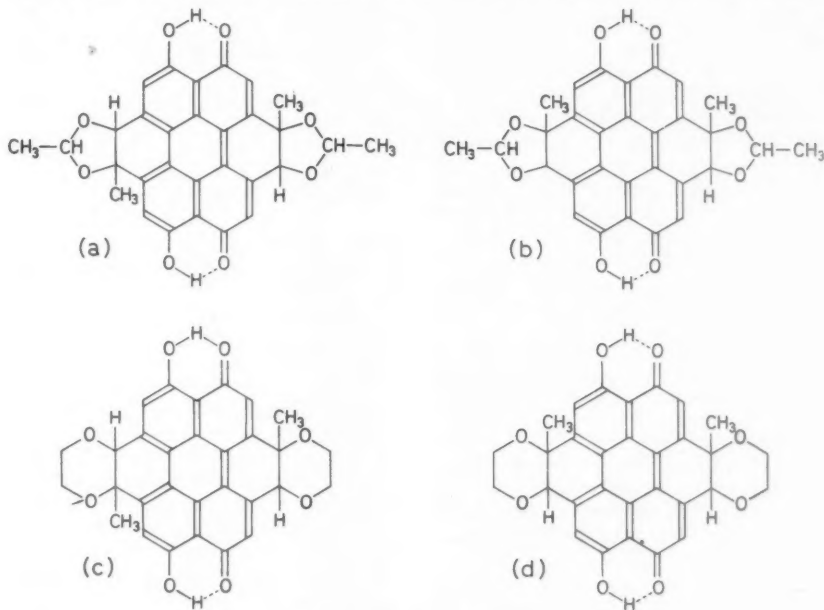


Figure 3. Alternative formulae for erythroaphin

the assumption that the heterocyclic rings are both *cis*-fused on to the rest of the molecule in erythroaphin fb, whereas in erythroaphin sl one of them has the less stable *trans*-fusion; if the angular hydrogen atom, which is activated by the quinonoid carbonyl group, is removed by the action of a base, it is replaced in the more stable configuration of the fb series¹⁰.

formulae (Figure 3a or 3b): all the major features to be expected in the hydrogen NMR spectrum on the basis of this formula are in fact observed, whereas a very different spectrum would be expected on the basis of the dioxan formulae (Figure 3c or 3d)¹². The yield of acetic acid in the Kuhn-Roth oxidation also receives a more natural explanation on the dioxolan formula. The striking difference in yield of

acetaldehyde on acid treatment between the two isomers can be explained on either formula if the primary reaction is stereospecific and is only possible with a cis-fusion of the rings, as would be

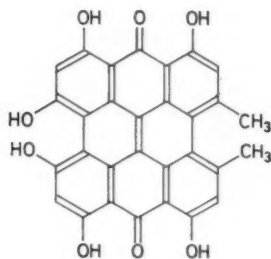


Figure 4. Hypericin

the case if it were a bimolecular elimination requiring a trans juxtaposition of the angular hydrogen and the oxygen atom of the heterocyclic ring attached to the adjacent carbon atom.

Thus although not definitely established the alternative formulae depicted in Figures 3a and 3b provide a good working hypothesis for the constitution of the erythroaphins, and further experiments designed to test it are in progress. The formula shown in Figure 3a is considered rather more likely, but the alternative (Figure 3b) is also possible; the stereo-chemical relations of the four methyl groups are unknown.

The Place of Erythroaphin among Natural Products

The polycyclic aromatic system of erythroaphin is surpassed in complexity among natural products only by hypericin (Figure 4), the quinonoid pigment of *Hypericum* spp (St. John's wort), and related plant pigments, the constitution of which has been elucidated by H. BROCKMANN and his collaborators¹³. Erythroaphin has many chemical and physical properties in common with hypericin and even shows some of its characteristic photo-dynamic activity.

The sparing solubility of erythroaphin in most solvents except chloroform and sulphuric acid is associated with the very strong internal hydrogen bonding, which also makes the four oxygens on the chromophore and the two sides of the aromatic system indistinguishable from the chemical point of view.

Xanthoaphin and Chrysoaphin

Chrysoaphin $C_{30}H_{24}O_9$ and xanthoaphin $C_{30}H_{26}O_{10}$ differ from erythroaphin by $1H_2O$ and $2H_2O$

respectively in their molecular formulae. As it is clear from their visible and ultra-violet spectra (see Figure 5) that they have quite different, less highly conjugated chromophores, the transformations

xanthoaphin \rightarrow chrysoaphin \rightarrow erythroaphin may be interpreted as a progressive aromatization of the molecule with elimination of water. Investigation of these pigments is hampered by the fact that this reaction is catalyzed by a great many reagents, which therefore yield the same products as would have been obtained from erythroaphin. Chromatography shows that whereas xanthoaphin

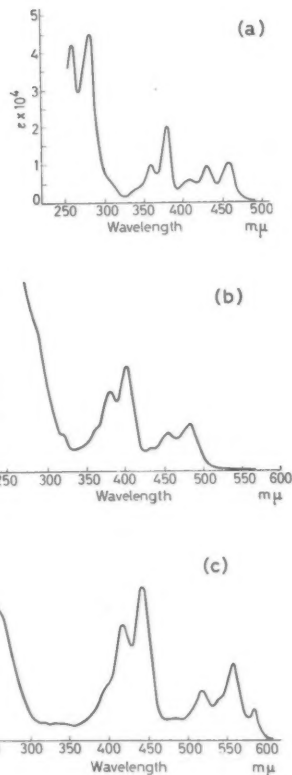


Figure 5. Light absorption in chloroform of (a) xanthoaphin sl, (b) chrysoaphin sl and (c) erythroaphin sl

fb and chrysoaphin fb are single substances, xanthoaphin sl consists of two and chrysoaphin sl of three components, which are probably stereoisomeric; all five can be converted to the unique erythroaphin sl¹⁴.

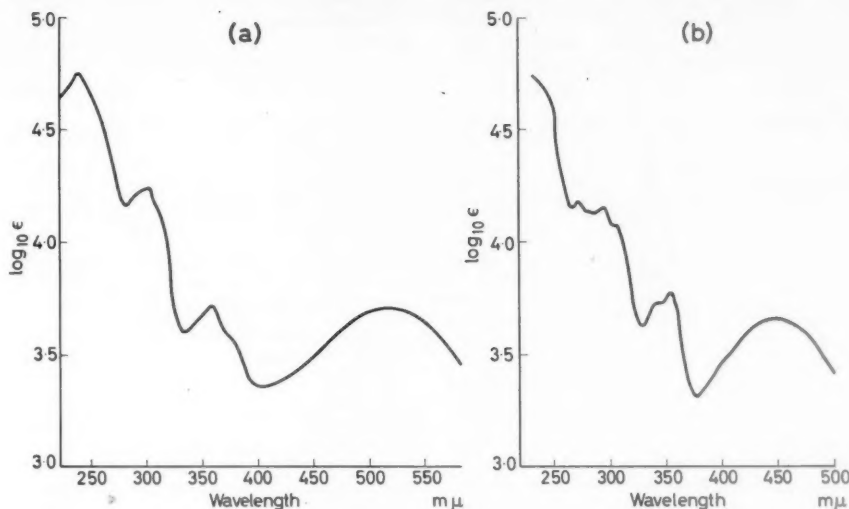


Figure 6. Light absorption of protoaphin fb (a) $\frac{7}{10}$ aqueous Na_2HPO_4 and (b) in 75 per cent alcohol

Protoaphin

Protoaphin, the pigment that is actually present in intact insects, differs radically from the other aphins. Unlike them it is very sparingly soluble in most non-hydroxylic solvents but somewhat soluble in water and freely soluble in mixtures of water and other solvents. It owes these properties to its glucosidic nature, which has been demonstrated by its hydrolysis by hot dilute acids to yield glucose¹⁵. In contrast to this process, which produces only a minute yield of erythroaphin among ill-defined amorphous products, the action of the enzyme present in the insects rapidly converts protoaphin to xanthoaphin in good yield. The enzymic reaction is accompanied by the uptake of one atom of oxygen per molecule and is evidently a much more complex process than the simple hydrolysis of a glucoside. Protoaphin fb has a pK of about 6 so that it exists chiefly as its deep red anion at neutral pH and is very largely responsible for the almost black colour of the living insects of many species. (Absorption curves for protoaphin fb are shown in Figure 6.) The free protoaphin is yellow and, unlike the anion, can be extracted from aqueous solution by butanol, which makes it possible to purify it by counter current distribution. Elementary analysis of protoaphin fb is difficult because of its hygroscopic nature but a molecular formula of $\text{C}_{28}\text{H}_{40}\text{O}_{18}$ (i.e. xanthoaphin + glucose $\text{C}_6\text{H}_{12}\text{O}_6$ + 2H) is the most likely¹².

It is not known whether protoaphin has any biological function in the insect or whether it is merely a waste product of its metabolism. An obvious possibility is that it acts as a respiratory pigment since it can be reduced to a leuco-compound which is readily reoxidized to protoaphin by aerial oxygen. Protoaphin occurs in solution in the blood of the aphid and is absent from the mycetomes. The raw material for its synthesis is probably the sugars absorbed by the aphid from the sap of its host plant, but nothing is known of the pathway by which this transformation occurs.

References

- ¹ SORBY, H. C. *Quart J. microscop. Sci.* (1871) **11**, 352
- ² BLOUNT, B. K. *J. chem. Soc.* (1936) 1034
- ³ HUMAN, J. P. E., JOHNSON, A. W., MACDONALD S. F. and TODD, A. R. *ibid* (1950) 477
- ⁴ DUEWELL, H. *et al. ibid* p. 485
- ⁵ — *et al. ibid* p. 3304
- ⁶ JOHNSON, A. W. *et al. ibid* (1951) 2633
- ⁷ BROWN, B. R. *et al. ibid* (1954) 107
- ⁸ CALDERBANK, A. and JOHNSON, A. W. and TODD, A. R. *ibid* p. 1285
- ⁹ BROWN, B. R. *et al. ibid* (1955) 954
- ¹⁰ — *et al. ibid* (1955) 959
- ¹¹ — *et al. ibid* (1955) 1144
- ¹² Unpublished work
- ¹³ BROCKMANN, H. *Fortschr. Chem. org. Naturst.* (1957) **14**, 141
- ¹⁴ JOHNSON, A. Q., TODD, A. R. and WATKINS, J. C. *J. chem. Soc.* (1956) 4091
- ¹⁵ BROWN, B. R. *et al. ibid* (1952) 4925

ELECTROSTATIC ACCELERATORS FOR NUCLEAR RESEARCH

D. R. CHICK AND S. E. HUNT

Research Laboratory, Associated Electrical Industries Ltd, Aldermaston Court, Aldermaston

Electrostatic generators developed from the early designs by Van de Graaff are used in medicine to produce fast electrons and thereby x-rays for diagnostic and therapeutic work. Electrostatic generators are also extensively used to accelerate positive charged particles; the beams of energetic nuclei thus provided remain an important tool in nuclear research. An increasing use of such generators is to provide a source of neutrons. Although, compared with the large circular accelerators built since the war, the energies are low, the authors point out that there are some unique advantages associated with machines of the Van de Graaff type for fundamental nuclear research. Some improvements and developments are discussed.

THE ELECTROSTATIC generator stems from the early design of Van de Graaff¹. The machine is widely used in medical science to produce fast electrons and consequently x-rays for diagnostic work and x-ray therapy, but is best known in the nuclear physics field as an accelerator of positive charged particles used in the investigations of fundamental nuclear physics problems.

Compared with large circular types of accelerators, or pulsed linear accelerators², the particle energies attainable are low, but for nuclear physics investigations the van de Graaff accelerator possesses several unique advantages which make it the most popular accelerator in nuclear physics laboratories throughout the world.

Operation of Van de Graaff Accelerators

The basic principles of operation are simple and are shown schematically in *Figure 1*. A high voltage terminal *T* is supported on an insulating column and charge is conveyed to it by a belt *B*, running between the lower and upper pulleys *P*, the lower being at earth potential and motor driven. Charge is sprayed on to the belt by means of a corona discharge from the row of fine needle points *S*, the voltage necessary to produce this being provided by a conventional rectifier unit, capable of producing voltages up to about 50 kV. Charge is removed from the belt at the upper terminal by means of a collector comb *C*, which may be similar to the spray comb and is situated inside the high voltage terminal.

The high voltage terminal continuously receives charge from the belt and consequently reaches a high positive potential which is either limited by voltage breakdown, or reaches an equilibrium value when the total current drain from the terminal is equal to that reaching it. In order to split up the

voltage along the length of the generator, the electrostatic column is built up of a series of conducting plates which are separated by insulators and tapped into a high value resistance chain running from the high voltage terminal to earth potential. Smooth hoops are fitted round the circumference of these so-called 'equipotential plates' which have holes to accommodate the belt and accelerator tube.

The positive particles to be accelerated are produced in an ion source situated inside the high voltage terminal and are repelled by the positive potential down an evacuated accelerator tube. They reach the bottom of the tube with an energy equal to their charge multiplied by the potential difference through which they have fallen.

Advantages of the Electrostatic Generator

It is seen, therefore, that a continuous beam of particles is available which have an energy stability determined almost entirely by the precision with which the positive terminal voltage can be kept constant. If the current drawn from the spray points is stabilized and the current leaving the ion source is constant, this stability is of the order of one per cent, but, as will be seen later, auxiliary stabilizing circuits can be employed to keep the beam energy constant to an accuracy of better than 0.05 per cent.

Further, it is possible to vary the energy of the beam in a controlled manner from a very low value to the maximum set by the machine geometry. The importance of this is easily seen when one considers the energy levels of the nuclei formed in excited states. It is well known that nuclei can exist only in certain states of quantized energy and properties of the nuclei when in these states are of fundamental importance to the study of the rather mysterious forces which hold the nucleus together. It is

important to determine precisely the excitation energies and also to make measurements to determine their spins, parities and modes of decay. In certain cases the states may be separated by only a few thousand electron volts of energy and may be extremely narrow; it is therefore necessary to use incident beams of extremely well defined energy in order to form the nuclei in the required state and to be able to vary this energy at will to excite different energy states in the same nucleus.

Figure 2 illustrates how an electrostatic generator meets this requirement. A target containing fluorine nuclei is irradiated by protons of extremely well defined energy. Protons are captured preferentially when their energy is just sufficient to form a new nucleus, neon 20, in one of its permissible excited states and since the states de-excite very rapidly by the emission of alpha particles accompanied by gamma rays, the maxima in the gamma ray yield curve correspond to excitation of the permissible energy levels in the Ne^{20} nucleus. In order to measure a yield curve of this type, the total energy

energy must be capable of being varied in a controlled manner in steps of fractions of one thousand electron volts. These facilities are found only in the electrostatic generator, and to a lesser extent in the Cockcroft-Walton accelerator. The latter, however, becomes uneconomic for energies above one million electron volts.

The electrostatic generator also has the advantage of being reasonably simple and reliable, so that research work can be carried out by a relatively small team with limited financial resources.

Electrostatic Generator as Neutron Source

This type of accelerator is also extremely useful as a source of neutrons. Accelerated deuterons striking targets rich in deuterium or tritium cause nuclear reactions in which neutrons are emitted. Since both reactions are exothermic, high energy neutrons can be produced using accelerated beams of quite low energy. For a fixed deuteron energy the neutron energy varies with the angle between the incident beam and the emitted neutrons, being higher in the direction of the deuteron beam and lower in the opposite direction. For example, if a 2 MeV deuteron beam is used to irradiate a deuterium target, 6 MeV neutrons are observed in the 'forward' direction and 1.9 MeV neutrons in the 'backward' direction. For a tritium target the energies are about 20 MeV and 12 MeV respectively. Neutrons of energies between those provided by these two reactions can be obtained by bombarding nitrogen targets with a deuteron beam and lower energy neutrons can be obtained by using a proton beam bombarding lithium or tritium targets.

Thus by a careful choice of target material and angle, reasonably mono-energetic neutron beams of almost any energy between a few thousand electron volts and twenty million electron volts can be obtained³. Such mono-energetic groups are not readily available from nuclear reactors and electrostatic generators have an additional advantage in that by suitably pulsing the charged particle beam, a pulsed source of neutrons can be obtained. This is a useful facility for experiments in which the decay times are to be measured.

The intensity of the neutron fluxes are, however, much lower than those obtainable from reactors, but a rather specialized type of electrostatic generator characterized by extremely high beam currents can be used to produce more intense sources of neutrons⁴.

Electrostatic generators also find application as injectors of charged particle beams into large circular accelerators, for which injection energies of between a few hundred thousand and several million electron volts are often required.

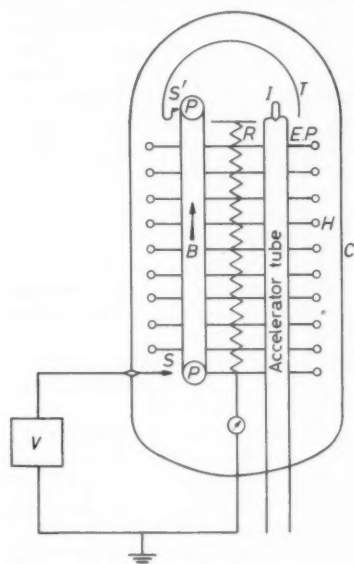
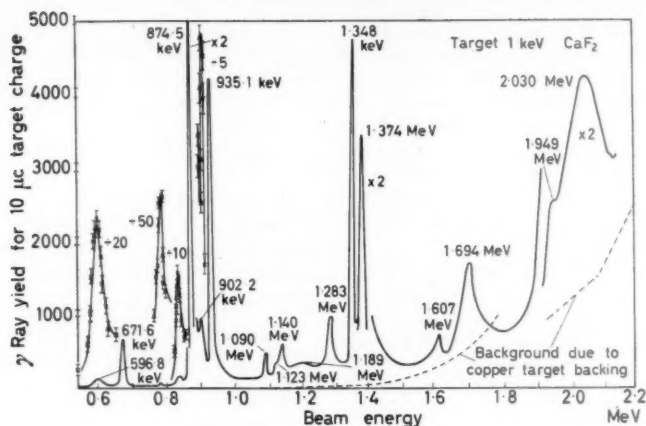


Figure 1. Schematic diagram of generator

- | | | | |
|------|-----------------------------------|---|-------------------|
| H | Charging belt | I | Ion source |
| T | High voltage terminal | V | Voltage supply |
| PP | Pulleys | C | Containing vessel |
| E.P. | Equipotential plates | R | Resistance chain |
| SS | Points producing corona discharge | H | Corona loops |

spread of the incident beam must be considerably less than one part in a thousand, and the mean

Figure 2. $F^{19}(p,xy)O^{16}$ yield curve

Methods of Energy Measurement and Control

The one per cent 'natural' stability of the electrostatic generator is not sufficiently high to investigate individual, closely-spaced levels. Auxiliary stabilizing circuits are required, but it is first necessary to measure the energy of the beam, which can be done either by electrostatic or magnetic deflection.

A particle of charge e , in falling through the potential difference V of the accelerator, acquires an energy Ve , and if it is subjected to a radial electric field by applying a potential difference X between two curved electrodes of separation d and geometric mean radius a , the relationship between V and X is

$$V = X \frac{a}{2d} \left(1 - \frac{eV}{2mc^2} \right).$$

Geometric dimensions a and d can be measured to extremely high precision, as can the deflecting voltages which are of the order of one hundredth those of the accelerator potential difference, so that the beam energy can be measured to a precision of the order of half a part in a thousand. The term in brackets is a relativistic correction which is small for energies of the order of a few million electron volts. Electrostatic analysers employing a deflecting angle of 90° have been made by several workers⁵⁻⁷. In general, the target currents obtainable at the output of these instruments are low (of the order of a micro-amp). By using a deflecting angle of $\pi/2\sqrt{2}$ radians much higher currents can be obtained without loss of precision or resolution, since a relatively wide parallel incident beam from the electrostatic accelerator can be focussed to pass through narrow

exit slits. Two such analysers have been constructed and used at this laboratory^{8,9} and a third one employing the principle has been used by R. BARJON¹⁰.

The main advantage of electrostatic analysers is that absolute measurements of beam energy can be made if care is taken in measuring the geometry of the analysers and the deflecting voltages accurately.

Magnetic types of deflectors are much more common. The energy of the deflected particle and the strength of the magnetic field H , are related by the well known formula

$$H\rho = c \sqrt{\frac{2mV}{e}}, \quad (\text{Gaussian units})$$

where ρ is the radius of curvature, and a parallel incident beam is focussed after deflection through an angle of 90° , which is extremely convenient, since most Van de Graaff accelerators are built vertically, whereas a horizontal beam is usually more convenient for nuclear physics investigations.

Magnetic deflectors have the additional advantage that problems of high voltage insulation in vacuum in the presence of beams of ionizing particles do not arise. In general, they are not used to measure the beam energy absolutely in view of the difficulty in measuring the average value of H along the trajectory of the particle, but are calibrated by observing nuclear events such as narrow (py) resonances or neutron thresholds, which have already been measured absolutely. The magnet exciter current or preferably the magnetic field at some reference point a small distance from the trajectory can then be related to these known energies. An important exception is the magnetic deflector used by F. BUMILLER and H. H. STAUB¹¹, where by careful

measurement of the field along the trajectory and of the reference point, an absolute accuracy of half a part in a thousand is claimed.

Having measured the beam energy, it is necessary to ensure that it remains constant within the limits set by the resolution of the electrostatic or magnetic beam deflector. This is done by taking signals from the exit slits of the instrument and using them as correcting signals. If for example the beam energy becomes too high, the radius of curvature of the beam will be larger than that defined by the geometry of the deflecting system and a large proportion of the beam will fall on the outer exit slit jaw. It is necessary for this to provide a signal which can be transmitted to the high voltage terminal of the machine in order to reduce the voltage and consequently the beam energy. Similarly, if the beam energy is too low, a signal from the inner slit must be used to increase the machine voltage and the mode of transmission of the signals to the high voltage terminal must be extremely rapid.

There are three well tried methods of conveying the correcting signal to the high voltage terminal. In the most common of these a set of needle points is mounted at a point fairly near to the high voltage terminal, such that a corona current flows from the terminal to the points. The electrical field gradient at the points is varied in a manner controlled by the error signal from the slits in such a way that when the machine voltage is too high more corona current is drawn, when it is too low the corona current is reduced¹². In this way, a variable current is drawn from the high voltage terminal in order to keep its potential constant.

Alternatively, a beam of electrons can be projected up the accelerator tube from an electron gun situated at the bottom of the tube, so that the electrons strike the tube anode plate. The potential of the grid of the gun is controlled by the error signals and consequently the intensity of the return electron beam can be varied in such a way as to keep the machine voltage constant¹³. This system has the advantage that the response is considerably faster, since the transit time of electrons in the accelerator tube is much less than the time for corona current to flow between the high voltage terminal and the corona points. It has, however, the disadvantage that on striking the top of the accelerator tube, the electrons produce x-rays, which, as will be seen later, are undesirable.

A third method consists of surrounding the Van de Graaff accelerator by a metal shield, the voltage of which is controlled by the error signals. When the voltage of the shield is varied, capacitive coupling between it and the high voltage terminal

induces variations in the voltage of the terminal, and by varying the voltage of the shield by means of the error signals, that of the high voltage terminal can be stabilized. Voltage variations of several tens of kilovolts are necessary on the shield, but the system has been found to work very satisfactorily¹⁴ and is particularly useful in maintaining the high voltage terminal constant if large pulsed beam currents are required from the accelerator.

Struggle for Higher Energies

In common with most other types of accelerator, it is desirable to obtain as high an energy as possible from a given size of machine. In the early days it is true that giant air-insulated machines were built in order to obtain high energies¹⁵, but the advantages of obtaining higher voltage gradients by enclosing the machines in gas at high pressures was quickly realized. The most commonly used gaseous dielectrics are mixtures of nitrogen with either about 20 per cent of carbon dioxide, or about 10 per cent of Freon (CCl_2F_2).

The choice of optimum pressure presents a difficulty in that the dielectric strengths of all gases increase with pressure, thus inhibiting corona loss and sparking, whilst one relies on just the former mechanism to charge the belt of the machine. In general, it is found that a pressure of about 200 lb/in² effects the best compromise for nitrogen-carbon dioxide mixtures and 150 lb/in² for nitrogen-Freon mixtures. At pressures in excess of this difficulty is often experienced in maintaining a steady corona condition at the belt spray combs and sparking can occur. This results in damage to the charging belt. Nevertheless, a very useful increase in voltage is obtained by running these machines in compressed gas. Air at normal pressure has a dielectric strength of about 30 keV per centimetre between plane electrodes, although severe corona sets in before this voltage gradient is reached. By using 10 per cent nitrogen-Freon mixtures in the Associated Electrical Industries electrostatic generator (Figure 3), terminal voltages of up to 5.5 million volts have been obtained across a nine-foot long stack in the absence of an accelerator tube; this has been calculated to be equal to 160 kV per centimetre across the main radial gaps. The variation in terminal voltage obtainable as a function of pressure for nitrogen and Freon-nitrogen mixtures is shown in Figure 4. This machine is described in detail elsewhere¹⁶.

An even stronger gas dielectrically is sulphur hexafluoride. This is extremely costly at present so that the cost of using pure sulphur hexafluoride would be prohibitive. However, it has recently

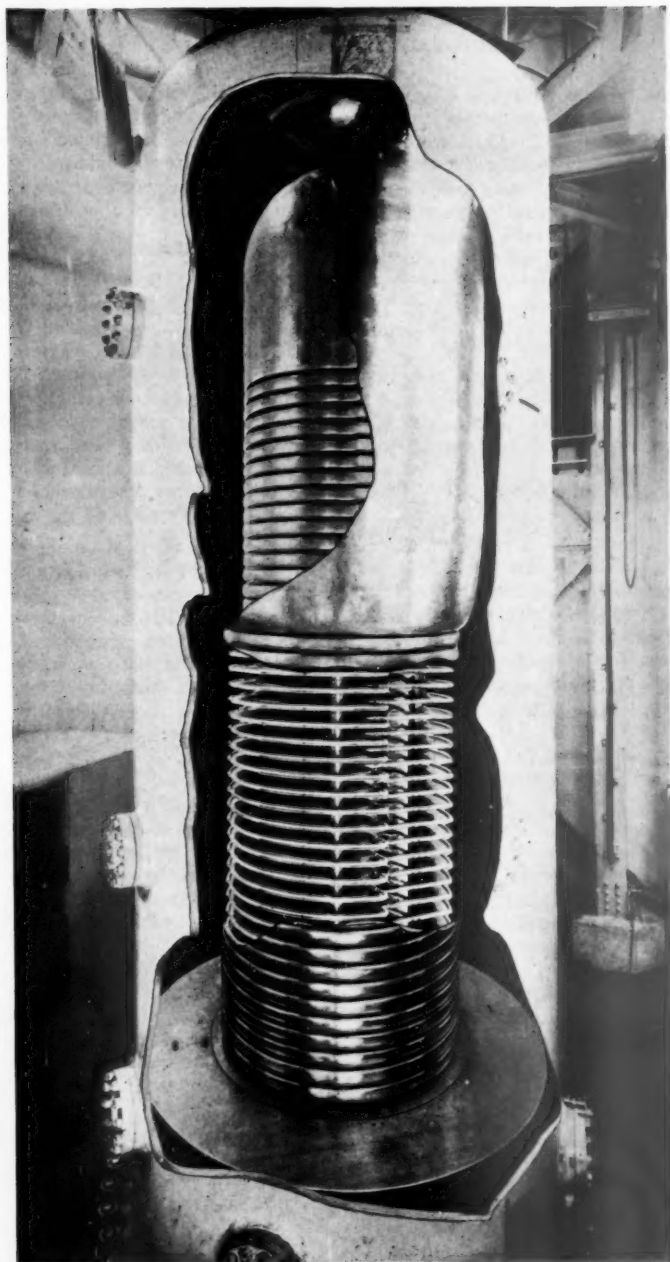


Figure 3. Electrostatic Accelerator installed at the Research Laboratory, A.E.I., Aldermaston, Berkshire

been found that very low concentrations of SF_6 in nitrogen-carbon dioxide mixtures considerably improved the dielectric strength¹⁷.

It will be observed in *Figure 3* that an additional electrode placed between the high voltage terminal and the earthed pressure vessel and held at an intermediate voltage can be used. This has the effect of splitting up the main radial gap and the increase in voltage attainable can be considerable, as shown by comparison of curves (b) and (c) of *Figure 4*. Other machines have been designed with two or more intermediate electrodes¹⁸. However, as will be seen later, the energy available is often limited by breakdown of the vacuum insulation in the accelerator tube and consequently the added complication of intermediate electrodes to improve the dielectric strength of the machine itself is often not justified. It will also be noted from *Figure 3* that great care is taken to ensure that all surfaces are smooth and well polished and that sharp corners are avoided in order to minimize loss due to corona and sparking. In the particular installation great care has been taken in the design of the corona hoops to ensure that the vertical and radial voltage gradients are optimized¹⁹.

In order to take full advantage of the improvement in the dielectric strength of the gas, similar improvements are called for in the voltage gradients which can be applied to the evacuated accelerator tube, the charging belt and the solid insulators supporting the stack. With the pressure of the gas adjusted to allow the correct type of corona discharge from the spray needles, it is found that discharges can take place along the length of the belt when the charge density on the belt surface reaches about one-tenth of that predicted theoretically and it is also found that these end-to-end discharges occur for lower charge densities as the vertical voltage gradient along the belt is increased. Experiments at this laboratory have illustrated that a fairly high proportion of the current originally sprayed on to the belt is not stripped off but circulates, increasing the charge density both on the up-going and down-going sides of the belt and thus increasing the probability of vertical sparking. Measurements of the charge density on the down-going side of the belt have shown that more complete stripping can be obtained by placing a positively charged plate near the inside of the belt opposite the stripper comb.

By suitably increasing the positive charge on this plate negative charges can be induced on the down-going side of the belt, increasing its effective charge carrying capacity before sparking occurs. A similar effect can be obtained by using a second

negatively charged set of spray combs at the top of the machine. On many machines wire gauze brushes touching or almost touching the belt are used to replace the more conventional needles of the stripping comb. End-to-end belt flashover is also thought to be due in part to an unequal distribution of charges across the belt width and many machines use metallic bars and shields near the belt surface in an attempt to spread the charge evenly over the belt width. Methods of charging and stripping the belt are in fact just about as numerous as are the machines themselves, each having its particular advantage for the chosen gas mixture, pressure and machine geometry. A practical upper limit to the amount of current transferred appears to be of the order of 20 to 30 microamps per inch width of belt using belt speeds of about 4000 feet per minute in the presence of a vertical voltage gradient of about 500 keV per foot length. The usual belt material is woven cotton impregnated with rubber and it is essential that belts are very carefully dried before use, otherwise they suffer electrical damage in a very short time.

The accelerator tube is normally evacuated continuously to a pressure of about 10^{-6} mm of mercury and consists of a high quality electrical porcelain or glass shell with metallic electrodes fitted internally. These are tapped into the resistance chain which runs from the high voltage terminal to earth and thus a smooth gradation of potential is provided along the length of the tube. In modern pressurized machines the behaviour of the accelerator tube usually imposes the limit to the energy obtainable. Carefully polished metal electrodes after conditioning can be shown to withstand voltage gradients of some 200 keV per centimetre at the surfaces under good vacuum conditions, yet it is often found that the mean gradient obtainable along a long accelerator tube is limited to some 15 kV per centimetre. Over these long lengths of the order of ten feet and in the presence of accelerated positive ion beams, the problem is no longer one of simple voltage breakdown between adjacent electrodes.

In most tube designs when the voltage gradient approaches the above figure, electron currents can be measured striking the top of the accelerator tube. Some of these electrons appear to be produced at the cathode end of the tube, due to ion and neutral particle bombardment in a manner suggested by R. L. FORTESCUE and W. K. MANSFIELD¹⁰ and are thus accelerated to almost the full machine energy. If at the anode of the tube they eject further positive ions, there is the possibility of a self-sustaining breakdown mechanism and indeed this type of

breakdown has been observed²⁰. The probability of such 'end-to-end' breakdown phenomena can be reduced by limiting the solid angle which each end of the tube subtends at the other. This is done by reducing the diameter of the apertures in the electrodes progressively to some point near the centre of the tube and is known as 'waisting'.

There is also very strong evidence that many electrons reach the top as a result of secondary multiplication processes at the surfaces of the electrodes and by specially shaping these to form geometric 'electron traps', retrograde electron current can be very appreciably reduced²¹. Retrograde electron currents can also be reduced by applying suppressing potentials near the anode and cathode of the tube and by applying reversed potentials over short lengths of tube²².

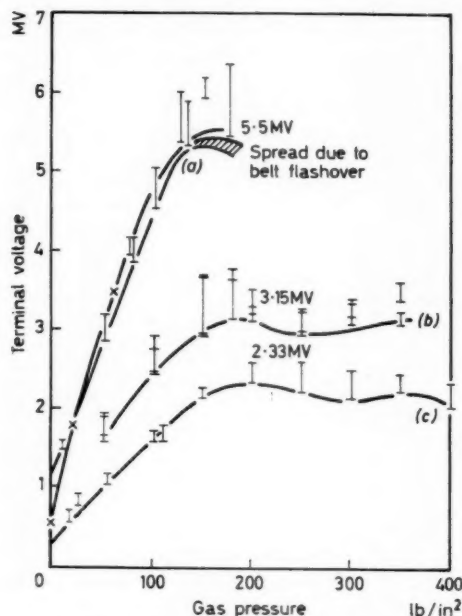


Figure 4. Voltage obtainable from the A.E.I. Electrostatic Generator as a function of pressure for nitrogen and a nitrogen-Freon mixture

The effect of retrograde electron currents is to produce x-rays on striking the anode of the accelerator tube. These in turn produce ionization currents in the high pressure gas between the high voltage terminal and the earthed pressure vessel. The Van de Graaff accelerator at A.E.I. Research Laboratory is perhaps unique in that currents reaching the

pressure vessel, currents leaving the high voltage terminal and those reaching the anode of the accelerator tube can all be measured²³.

It has been found that current drains of the order of a hundred microamps can under certain circumstances flow radially across the gas from the high voltage terminal to the pressure vessel. These can be produced by tube electron currents of ten or twenty microamperes reaching the anode. This unwanted current drain on the machine may limit the voltage attainable before actual voltage breakdown of the end-to-end type occurs in the tube²¹.

In early pressurized machines the communication system used to control the variables situated in the high voltage terminal, such as the ion source settings, etc., were strings and insulating rods and sparking along the length of these was a frequent source of trouble. This has been overcome in the Van de Graaff machine at this laboratory by the use of a modulated light beam, of which the upward going beam acts as a control system and a downward going beam conveys the information on the settings of the relevant parameters. The scheme has the advantage of lending itself to remote control and that channels can be switched to measure different parameters such as retrograde electron current, etc²³, apart from the obvious one that no physical links are involved from the high voltage terminal to earth.

Future Developments

A large number of Van de Graaff machines now exist in which voltage gradients of about 15 kV per centimetre are obtained along the length of the stack and tube. The largest of these is the machine at Massachusetts Institute of Technology, where a beam energy of eight million electron volts is achieved using an 18-foot long accelerator tube²⁴. If higher energies are required, the tendency now is to build the so-called 'Tandem' machines.

In principle, these are similar to the simple electrostatic accelerators. A positive voltage terminal situated at the centre of the machine is charged by the current from an endless moving belt. The accelerator tube passes through the terminal and protrudes at each end of the pressure vessel, so that both ion source and target end of the beam tube are at earth potential.

The ion source conditions are arranged to produce protons to which an excess of electrons are attached. These extracted ions carrying a net negative charge are separated from the electrons, which are also drawn from the source, by means of a magnetic field. The desired heavy ions in this negatively charged state are accelerated towards the positive central

terminal. Here they pass through a region of higher gas pressure or through a thin foil which strips off the attached electrons. The resultant positive particles are now repelled by the positive terminal



Figure 5. The tandem electrostatic accelerator at A.W.R.E. Aldermaston. This is an inside view of the pressure vessel of this accelerator showing the column supporting the high voltage terminal. It is designed for the production of proton beams of 12 million electron volts energy. The machine is contained in a pressure vessel 45 feet high by 9 feet internal diameter and works at a tank pressure of 200 pounds per square inch. It will also be possible to accelerate heavy ions which is a unique field of research with this type of machine

and further accelerated down the second half of the accelerator tube to reach the target with energy equal to their charge multiplied by twice the terminal potential.

This system retains to a high degree the advantages of the simple electrostatic generator; the beam is homogeneous in energy and can be stabilized to a high degree of accuracy and has the additional advantage that the ion source and associated supplies are easily accessible without removal of the pressure vessel. For a given final energy the radial clearance from high voltage terminal to pressure vessel is much reduced because of the double acceleration, resulting in an appreciable economy in space and the cost of the pressure vessel.

Appreciable losses of beam are involved in the electron pick-up and stripping mechanism, but since the cross sections for nuclear events are much larger at these higher beam energies, this loss is not quite so important as would appear at first sight.

The first machine of this type was reported to be in operation at Chalk River last year²⁶; two similar machines are now in operation at the Atomic Weapons Research Establishment (see Figure 5) and the Atomic Energy Research Establishment. These machines are designed to produce positive particles of energies up to twelve million electron volts.

The authors wish to thank Dr T. E. Allibone, F.R.S., Director of the Laboratory, for permission to publish this paper.

References

- ¹ VAN DE GRAAFF, R. J. *Phys. Rev.* (1931) **38**, 1919
- ² CHICK, D. R. and MILLER, C. W. *Brit. Communications Electronics* (Oct. and Nov. 1956) 539, 596
- ³ HANSEN, A. O., TASCHEK, R. F. and WILLIAMS, J. H. *Rev. mod. Phys.* (1949) **21**, 635
- ⁴ FELICI, N. J. *Brit. J. Appl. Physics, Suppl.* 2, (1953) S 62
- ⁵ HERB, R. G., SNOWDEN, S. C. and SALA, O. *Phys. Rev.* (1949) **75**, 246
- ⁶ FOWLER, W. A., LAURITSEN, C. C. and LAURITSEN, T. *Rev. Sci. Instrum.* (1947) **18**, 818
- ⁷ BONDELID, R. O. and KENNEDY, C. A. Naval Research Lab. (Washington) Report No. 5083 (1958)
- ⁸ HUNT, S. E. *Proc. phys. Soc. A.* (1952) **65**, 982
- ⁹ HUNT, S. E. et al. *Proc. Instn. elect. Engrs.* (1956) **103B**, No. 8, 146
- ¹⁰ BARJON, R. *Thèse du Doctorat* (Paris, 1954)
- ¹¹ BUMILLER, F., STAUB, H. H. and WEAVER, H. E. *Helv. Phys. Acta* (1956) **29**, 83
- ¹² ASHBY, R. M. and HANSEN, A. O. *Rev. sci. Instrum.* (1942) **13**, 128
- ¹³ CHICK, D. R. and INALL, E. K. *Brit. J. appl. Phys. Suppl.* 2 (1953), 61
- ¹⁴ GABET, A. and TAIEB, M. *L'Onde Electrique* (Nov. 1955), **35**, 1033
- ¹⁵ VAN ATTA, L. C. et al. *Phys. Rev.* (1936) **49**, 761
- ¹⁶ CHICK, D. R. and PETRIE, D. P. R. *Proc. Inst. elect. Engrs.* (1956) **103B**, 132
- ¹⁷ PIROVER, L. I. and TUBAER, V. M. *Zur. Tech. Fiz.* (1958) **28**, No. 7, 1538
also BUECHNER, W. W., private communication
- ¹⁸ FORTESCUE, R. L. and HALL, P. D. *Proc. Instn. Elect. Engrs.* (1949) **96**, Part I, 77
- ¹⁹ MANSFIELD, W. K. and FORTESCUE, R. L. *Brit. J. appl. Phys.* (1957) **8**, 73
- ²⁰ HARRIS, D. J. *Ph.D. Thesis* (London, 1953)
- ²¹ CHICK, D. R. et al. *Nuclear Instrum. & Methods.* (In the press)
- ²² CHICK, D. R. and FIRTH, K. *J. Sci. Instrum.* (1953) **30**, 117
- ²³ MILLAR, B. M. *Electronic Engng.* (Nov. and Dec. 1953) 446, 506
- ²⁴ M.I.T. *Progress Report*, p. 93 (May 31, 1952), also BUECHNER, private communication
- ²⁵ GOVE, H. E. et al. *Phys. Rev.* (1958) **1**, 251

GAS LUBRICATION

J. A. COLE

*University of Western Australia, Nedlands, Western Australia**

Most machine bearings are lubricated with petroleum lubricants, but in some applications there are advantages to be gained from the use of gaseous lubricants, usually air. This article describes the principles and development of gas lubrication, and surveys some current applications.

AT FIRST sight, gas lubrication appears to be a new branch of technology, a product of the nuclear age which it serves. Gas bearings are now to be found in high speed grinding spindles, ultra-centrifuges, gyroscopes, gas circulators for nuclear reactors, high speed textile spindles, dynamometers and electric motors. Admittedly many of these applications are still in the experimental stage, but at least two British firms are advertising a gas lubricated pump suitable for circulating gases at temperatures up to 450°C (see *Figure 1*), while a proprietary air driven and air lubricated grinding head for speeds up to 100,000 revolutions per minute has been available for several years.

The intense interest and activity is certainly new, but as is often the case the first steps were taken many years ago, more or less concurrently with the investigations on the basic principles of film lubrication. But before tracing the historical background and dealing with the problems being investigated today, it must be considered what is involved in the gas lubrication of plain bearings.

Film Lubrication

The term film lubrication—sometimes thick film lubrication—is used to describe that desirable state of affairs in a plain bearing where the opposing bearing surfaces are completely separated by an intervening layer of lubricant of minimum thickness between 0.00004 in. and 0.001 in. Under these conditions of no solid contact between the high spots of the bearing surfaces, boundary friction is replaced by fluid friction, the friction coefficient falls from 0.3 or more to as low as 0.001 and mechanical wear is entirely absent.

The important physical property of the lubricant in film lubrication is its viscosity. (The property often described as 'oiliness' is a separate issue, and is only significant when incipient contact

occurs between the bearing surfaces.) Oils are normally used because their high viscosity gives adequate—and sometimes more than adequate—load capacity, but as speeds rise an increasing amount of heat is produced in shearing the oil film and eventually it becomes necessary to use a forced feed system to supply an excess of oil to the bearing for cooling purposes.

Current mechanical engineering designs often use bearings under conditions of high speeds and temperatures, with problems of lubricant stability, contamination and sealing. Gas bearing then becomes attractive. Gases have very low viscosities, about one thousandth of that of a conventional lubricant, so that the friction power loss, which may be excessive with petroleum lubricants at high speeds, is low and external cooling unnecessary. Gas viscosity also rises with increasing temperature whereas liquid viscosity is lowered, and most gases are chemically stable at temperatures well above the decomposition point of conventional lubricants. Needless to say, there are limitations: gas lubrication is suitable primarily for lightly loaded bearings and is therefore unlikely to displace oil lubrication in present mechanical equipment except where high speeds and light loads are involved. However, work is proceeding in both Britain and the United States on steam lubricated bearings for steam turbines, for example.

Types of Film Lubrication

Film lubrication in general, and gas lubrication in particular, depend on the generation of sufficient fluid pressure in the thin layer of lubricant between the opposing bearing surfaces to keep them apart under the applied load, and this pressure generation can be achieved in two ways.

A self-acting or hydrodynamically lubricated bearing generates the film pressure automatically, working on the principle of the 'fluid wedge', (*Figures 2 and 3*). The bearing surfaces are set,

* Formerly Lubrication and Wear Division, National Engineering Laboratory, East Kilbride

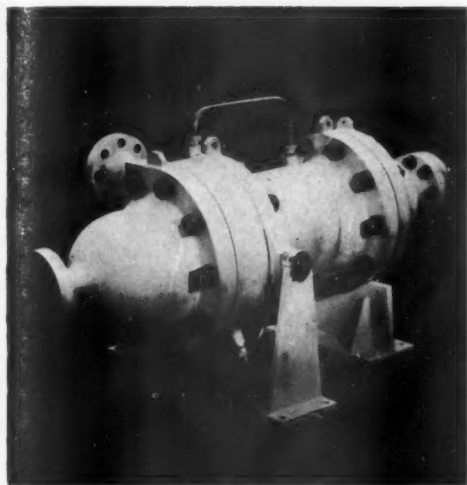


Figure 1. (LEFT) Circulator using self-acting gas bearings and (RIGHT) its rotor
(by courtesy of Ruston & Hornsby Ltd)

or adjust themselves, to form a wedge shaped space converging in the direction of fluid motion. The relative motion of the bearing surfaces drags lubricant into the restriction by viscous action, building up pressure sufficient to carry the load. The beauty of the system is its self-adjusting nature over a wide range of conditions. Under certain circumstances (outside the scope of this article) thermal expansion of the lubricant can lead to an equivalent effect with parallel bearing surfaces, the so-called 'thermal wedge'. This fluid wedge principle is most effective with lubricants of high viscosity, and gases are at a double disadvantage in that low viscosity and compressibility reduce the load capacity. The main deficiency of hydrodynamic lubrication is that it fails under starting and stopping conditions, as the relative motion is then insufficient to maintain an effective film.

An externally pressurized or hydrostatically lubricated bearing uses an external source, such as a pump, to supply lubricant to the film space at a sufficiently high pressure to prevent contact (see Figure 4). Thus film lubrication is effective irrespective of whether the bearing surfaces are in relative motion or not. Viscosity remains a significant property of the lubricant, since it governs the rate of escape from the lubricating film, and hence the capacity of the external pumping system; compressibility is sometimes of minor importance here. In practice, both mechanisms may be called into

play—a heavily loaded bearing may be started under hydrostatic conditions to bridge the interval before hydrodynamic operation is possible. A high speed hydrostatic bearing may derive some load capacity from hydrodynamic effects. (The inappropriate prefix 'hydro-' seems to be too well established, and rival terms too clumsy, to hope for a neater nomenclature. One American writer uses the term 'pneumodynamic' but this is hardly general enough.)

The Development of Self-Acting Gas Bearings

The first suggestion that air could be used as a lubricant is commonly attributed to G. A. Hirn¹ although further research may show that Leonardo da Vinci had ideas on the subject. Hirn is quoted² as stating:

'There is a remarkable difference between the friction of two pieces (*sic*) on each other when the surfaces are dry and when the surfaces are lubricated by some unctuous matter or air . . .'

and that

'air under certain circumstances and in sufficient quantity becomes the best lubricator, the coefficient then falling to 1/10,000'.

Others had already commented on the lubricating properties of the thin air film trapped between two surface plates³ in 1885 when Beauchamp Tower published the results of his classical experiments on the friction and lubrication (with oil) of railroad-

type journal bearings⁴. These experiments laid the foundation of the subject of film lubrication, demonstrating that substantial loads could be carried by plain bearings with friction coefficients as low as 0.0014 provided a copious supply of oil was maintained by means of an oil bath. Tower produced direct evidence that the loads were supported by pressures generated in the thin layer of oil between the bush and the journal.

Lord Rayleigh speculated on the hydrodynamical basis for this phenomenon and stated that

'it would appear to be necessary that the layer should be thicker on the ingoing than on the outgoing side'⁵.

He also predicted that Professor Stokes would provide an explanation for the results, but it was Osborne Reynolds who produced a comprehensive mathematical theory of film lubrication⁶ which explained the important features of Tower's experiments and confirmed Rayleigh's concept of the fluid wedge. The essential assumptions of the theory were that the lubricant was postulated to be an incompressible liquid of constant viscosity, the lubricant layer was taken to be thin and wedge shaped in profile, and the fluid forces were taken to be predominantly due to viscous flow, inertia effects being negligible. Reynolds' work provided the basis for all mathematical analyses of film lubrication, and the theory has since been greatly developed, making use of more advanced mathematical methods and catering for an ever wider range of physical conditions, now including the effect of compressibility.

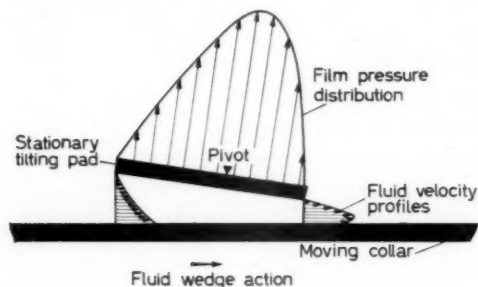


Figure 2. Fluid wedge action—the film thickness is greatly exaggerated

The work of Tower and Reynolds led to the development, at the turn of the century, of the tilting-pad thrust bearing, a plane slider bearing working on the fluid wedge principle. This was a practical development of the greatest importance, since it enabled the bulky and inefficient multicollar thrust bearing then used in steamships to take the end thrust from screw propellers to be replaced by

the far more compact and efficient tilting-pad single collar thrust bearing. The invention was made independently and more or less simultaneously by A. G. M. Michell in Australia and A. Kingsbury in the United States.

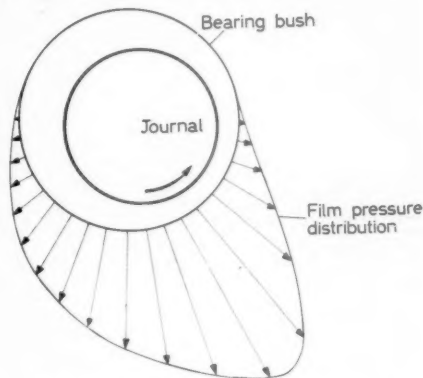


Figure 3. Fluid wedge action in a journal bearing—the film thickness is greatly exaggerated

Michell's work on a practical thrust bearing was coupled with an important development of the Reynolds theory enabling the bearing of finite width to be analysed. Kingsbury had been led to his conception of the tilting-pad thrust bearing by some chance observations of the lubricating properties of the air film between a cylinder and a closely fitting piston⁷. These observations he later related to Reynolds' work, and in 1897 he published the results of the first experimental investigation of the air lubricated journal bearing⁸. Kingsbury's air bearing ran at 4400 revolutions per minute and supported a load of one and one-third pounds per square inch of projected area, and he evidently regarded it as having no practical significance. In the 1922 catalogue of the Kingsbury Machine Co., the operation of the Kingsbury thrust bearing is illustrated by means of an air lubricated model, but it is stated that

'air is too thin a lubricant for ordinary use, although it serves the purpose of this demonstration'.

Harrison⁹ was the first to consider the theoretical aspects of gas lubrication, extending the Reynolds theory to deal with a lubricant which was compressible as well as viscous. The complication made the mathematics more intractable and a general solution was not obtained, but the theory explained in part the experimental results of Kingsbury. In 1921, Stone, another Australian, carried out detailed experiments on an air lubricated thrust

bearing in order to verify Michell's theoretical analysis, compressibility effects being regarded as small¹⁰. However, the practical possibilities of gas lubrication were not really examined until the second World War, although some *ad hoc* work had been done in the 1930s on air bearings for high speed laboratory centrifuges. The first published account of an engineering application of gas lubrication came from Germany, where H. DRESCHER¹¹ described electric motors fitted with gas lubricated plain bearings. Drescher used light loads on large bearings, about 4 lb/sq.in. of projected bearing area, running at conventional speeds (3000 rev/min). Earlier, patents for gas lubricated bearings had been taken out in various countries, but there seems to be little published information on successful operating experience.

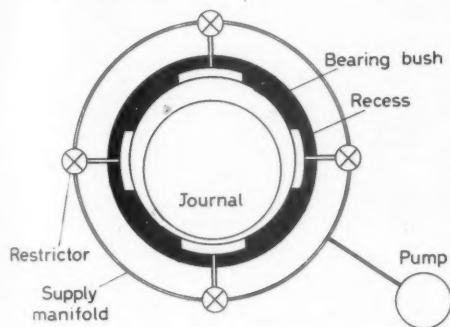


Figure 4. Externally pressurized journal bearing (diagrammatic)

Some interesting results were obtained by an American, O. C. BREWSTER, who developed the use of 'side air', or hydrostatic preloading, to suppress vibration in high speed hydrodynamic gas bearings¹². An intriguing feature of this work in these days of lavish research facilities and expenditure is that it was conducted as a hobby by Brewster, the research laboratory being a tiny workshop in the basement of a drug store in the small Connecticut town of Litchfield. The next comprehensive experimental investigation was reported from the Atomic Energy Research Establishment at Harwell¹³ which described a laboratory investigation of the performance of lightly loaded gas bearings ($1\frac{1}{2}$ to 13 lb/sq.in.) running at speeds up to 15,000 rev/min, and also gave details of two reactor applications, one a carbon dioxide circulating pump and the other a pump for radioactive bismuth. These authors made a point of using wider bearings than those employed in normal practice, with a diaphragm mounting to overcome difficulties of alignment.

More recently, results of experimental investigations in Britain, Russia and the United States, and of theoretical work in Roumania, Japan and the United States have become available¹⁴. Empirical design rules are beginning to emerge: work in Britain and America had indicated that a load capacity of about one pound per square inch per 1000 rev/min can be achieved under ideal conditions but, as will be shown later, there is still scope for a great deal of research and development work.

Externally Pressurized Gas Bearings

Hydrostatic lubrication is generally a rather more *ad hoc* subject than hydrodynamic lubrication, and its early historical development, expressed in patent applications and in incidental descriptions of equipment rather than in formal research papers, is correspondingly more difficult to trace.

D. D. FULLER¹⁵ considers an exhibit at the Paris Industrial Exposition of 1878 to be probably the earliest recorded application of hydrostatic lubrication: oil supplied under pressure provided a low friction support for a large stationary mass which could then be freely moved about on a base plate. A textbook method for designing journal bearings with hydrostatic oil lift was given¹⁶ as early as 1925 and this device for overcoming starting difficulties with heavily loaded hydrodynamic bearings was often used, although it was not until 1947 that a comprehensive analytical treatment of the principles and practice of hydrostatic lubrication was given.¹⁶

Patent for externally pressurized air bearings were taken out as early as 1920, and descriptions of laboratory ultracentrifuges or of high speed spinners for rotating mirrors using this type of bearing appeared during the succeeding twenty years, what seems to be first being by E. Henriot and E. Huguenard¹⁷. These investigators described an air driven, air supported spinner which was said to have reached a speed of 660,000 rev/min. However, as with self-acting gas bearings, it was not until World War II that research and development really began to flourish, and a considerable amount of rather diverse information on the design of hydrostatic air bearings is now available from papers published in France, Germany, Britain, Japan and the United States.

Research on Self-Acting Gas Bearings

Current research in this field is concerned with three broad topics, the fluid mechanics of the gas lubrication process, bearing materials and bearing stability. (An air bearing test rig is shown in Figure 5.)

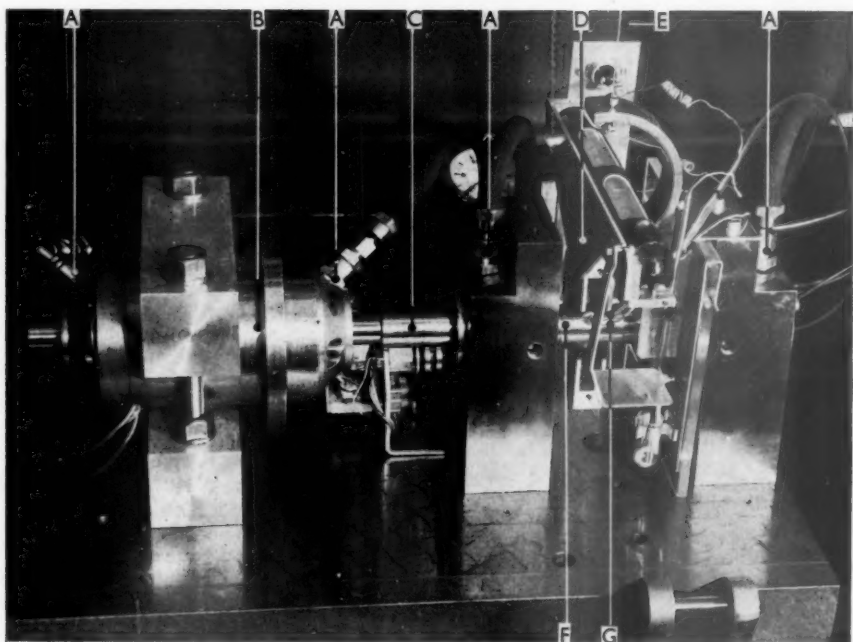


Figure 5. Air bearing test rig (by courtesy of the Institution of Mechanical Engineers)

- | | |
|--|--|
| A Externally pressurized main bearings | E Loading cable |
| B High frequency induction motor | F Test bearing |
| C Test shaft | G Film pressure tapping and thermocouple |
| D Friction torque arm | |

A comprehensive understanding of the fluid mechanics of gas lubrication is desirable if rational design procedures are to be worked out, although the intrinsic interest of a new branch of fluid mechanics is undoubtedly responsible for at least some of the research activity here. Conventional film lubrication theory postulates the laminar flow of a viscous incompressible fluid in the thin bearing film and calls for fairly involved mathematics to obtain solutions even for highly idealized situations. However, progress is now being made in the analysis of bearings lubricated with compressible fluids, with some surprising results. Usually, when a compressible fluid flows under conditions producing only small variations of pressure, it can be adequately described by incompressible flow theory, and this may be true for a Mach number—ratio of fluid velocity to local velocity of sound—as high as 0.3. This is not the case in gas bearings, where Mach numbers are small and yet film pressure variations may be very large.

The most curious implications of gas lubrication theory are that even under light load conditions

where pressure variations are quite small, the flow can still not be regarded as incompressible, and that under heavy loads the viscosity is not an important factor in load capacity. Both these points require experimental confirmation; indeed the basic question of whether the conditions in the film are effectively isothermal or adiabatic has yet to be finally settled. So far, only steady load conditions have been considered and the whole field of dynamic loading, including loading due to rotor unbalance which becomes very important at high speeds, awaits theoretical and experimental investigation.

Bearing materials present a problem because of starting and stopping conditions when the hydrodynamic film is inoperative and contact occurs. Accidental contact may also occur during running as a result of bearing misalignment, bearing vibration or contamination with particles of dust. Most gases, unlike petroleum lubricants, possess little or no boundary lubricating properties, and conventional bearing materials are therefore unsatisfactory except for insignificant loads. Plastics are unsuitable due to their poor dimensional stability and low

thermal conductivity, although bearing materials impregnated with P.T.F.E. offer some promise. Suggestions have been made that the difficulty of starting should be overcome by using either a hydrostatic lubrication or a liquid lubricant which boils off as the bearing accelerates to full speed.

Current practice is to use conservative bearing loads, although this may fail to exploit the full load capacity of the gas film, and hard bearing materials with a high surface finish, thus minimizing the tendency for surface welding. Chromium plated journals running on nickel alloy bushes have been used, and ceramics are a possibility. Bearing clearances near to or slightly less than the normal value of 1/1000 times the diameter seem to be favoured, but high precision is required as hard bearing materials do not allow running-in to take place in order to compensate for errors of geometry. Similarly, these materials do not possess one of the valuable properties of bearing white metals, that of being able to embed any foreign particles trapped in the film space, so that gaseous lubricants must be kept reasonably dust-free.

At very high temperatures, in the region of 500°C it may be possible to make use of the good boundary lubrication properties of reactive gases such as sulphur hexafluoride in conjunction with certain alloys, but while bearings for high temperature operation remain an urgent problem in the military field, not much information is available.

Bearing stability is concerned with hydrodynamic film whirl, a problem liable to be met with all lightly loaded high speed hydrodynamic bearings. It manifests itself as a vibration of the journal in the bearing clearance space with a frequency equal to half the running speed, and may be complicated by shaft elasticity if the machine is running at or above twice its critical speed. The vibration is usually undesirable, and if severe it leads to contact and bearing damage even though the nominal bearing load is small. The usual remedies are the use of higher loads, the application of hydrostatic preloading, or the use of stabilizing pockets in the bearing surface, but more experimental information on conditions producing whirl is needed.

Vapour lubrication, as distinct from gas lubrication, has received little attention, although it has been reported that air lubricated bearings do give rise to some water condensation in the clearance space due to water vapour in the air (Figure 6). Gas lubrication is an attractive possibility for refrigeration plant in order to eliminate sealing problems, especially since some refrigerants have good boundary lubricating properties, so that this application will probably receive attention in the future.

Externally Pressurized Gas Bearings

The basic principle of hydrostatic lubrication has already been explained, and also its particular advantage of being effective with or without bearing rotation. There are, therefore, two fairly distinct fields of application. A hydrostatic bearing may be used in non-rotating (or slowly rotating) equipment in order to provide a low friction support, for example, as a support-bearing in a dynamometer, or in a balance for weighing forces in a wind tunnel, a Couette viscometer, or a bearing test machine. A classical example is the mounting for the Mount Palomar telescope¹⁵. In such cases, the frictional resistance is vanishingly small, usually far lower than that offered by a ball bearing, so that high sensitivity and steadiness of motion are obtained. Of course, this is only achieved at the expense of the power used in the pumps supplying the bearing,

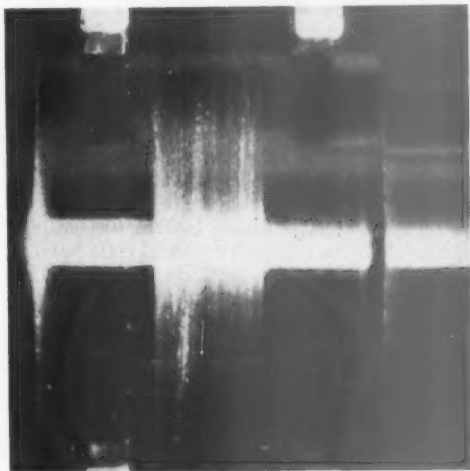


Figure 6. Water condensation in an air-bearing (by courtesy of the Institution of Mechanical Engineers)

so that the overall energy balance may not be very favourable. In general, oil is to be preferred in such applications in the interests of maximum load capacity and minimum pumping power. The use of air, steam or other gases involves a greater flow rate, but can be desirable, in some cases, since it gives cleaner operation and can avoid the necessity for lubricant recovery.

Hydrostatic lubrication may also be used in rotating bearings. Thus load capacity can be kept constant over a wide range of speeds from a state of rest, with controlled positioning of centre of rotation. The shearing of the oil film in a high speed

rotating hydrostatic bearing may give excessive friction loss and temperature rise, whereas low friction running is achieved by changing to gas lubrication. With a gas, hydrostatic instead of hydrodynamic operation may be essential to obtain adequate load capacity. The low torques and loads required for small grinding wheels rotating at speeds near 100,000 rev/min provide an example and pneumatic units are often used for this application, with an air turbine drive and with externally pressurized air bearings.

The disadvantages of externally pressurized gas bearings compared with their oil lubricated counterparts lie principally in the higher consumption of power in the external pump system and in the greater risk of vibration due to the high compressibility of the working fluid. To reduce the pump capacity required, smaller clearances are used, calling for good surface finish and geometric accuracy. Good filtration of the gas supply to exclude dust is then essential, and there have been cases of hydrostatic air bearings becoming quite inoperative due to small quantities of water or oil in the air supply clogging the fine clearance space.

The question of stability arises in several ways and depends on the form and arrangement of the gas admission channels in the bearing. In the simplest type the gas is supplied under pressure to one or more recesses in the bearing surface, and leaks away to the outside by way of fine clearance lands. If these recesses are large, the gas within them and the load form a spring-mass system whose natural frequency may give rise to vibration and damage, especially if it coincides with some disturbing force. At high speeds, hydrodynamic whirl may interact with recess compressibility effects, and failures at the National Engineering Laboratory have been attributed to this. Badly positioned recesses may give inadequate resistance to misalignment. The general conclusion seems to be that recesses should be shallower than in oil-lubricated bearings, and should be eliminated in favour of plain radial admission holes where the lower load capacity obtained with holes alone can be accepted.

To promote stability and to economize in pump capacity, it is desirable to use flow restrictors in series with the gas admission holes or recesses (Figure 4). These may take the form of orifices or of capillaries, often adjustable. When for some reason the film thickness at a particular supply hole increases, the increased outflow of gas gives rise to a larger pressure drop across the restrictor, so that the gas pressure available in the film is reduced. The lower load capacity at this point then tends to restore the film thickness to its original value,

giving a self-regulating mechanism. In addition, the throttling action of restrictor prevents excessive demands being placed upon the external pump. Sometimes the restrictors are not left to operate automatically, but are individually and manually adjusted to obtain maximum load capacity.

One particular form of bearing deserves mention in view of its simplicity. This makes use of a porous (sintered) metal bush of the type normally used as an oil impregnated, self-lubricated bearing. Used dry and supplied with air under pressure, it forms a useful self-regulating hydrostatic bearing, and can be mounted in rubber O-rings to serve as anti-vibration mounts and as seals. In this form, the bearing was first described by A. G. MONTGOMERY and F. STERRY¹⁸ but porous bushes are said to have been used in this manner at an earlier date in France and in the United States.

Research on Externally Pressurized Gas Bearings

Research and development work in the field of hydrostatic gas bearings is perhaps more active in the case of hydrodynamic bearings, partly because of the wider range of configurations which can be employed. Load capacity, gas consumption and stability under steady load conditions are normally studied. Load capacity can usefully be expressed in terms of the ratio load pressure to supply pressure, and at first sight it is surprising that this figure seldom approaches unity. The use of an appropriate distribution of large area recesses which ensure that the film pressure is approximately equal to the supply pressure and the provision of a large reserve of pump capacity can produce high ratios, but in general this is inefficient in terms of pumping energy. A ratio of between one-quarter and one-third is usually the most that is practicable, and many investigators have obtained far lower values. Most journal bearing designs involve a symmetrical array of pressure holes or recesses even when the load is unidirectional, and at the National Engineering Laboratory it has been found advantageous to supply air mainly on the loaded side.

Load capacity and friction in terms of pumping energy have been less studied. For a rotating bearing, as film thickness rises, shearing friction decreases but gas flow and hence pumping energy increase, so that an optimum film thickness exists but there is no general rule to use in design. Other topics requiring further study are the choice of optimum gas supply arrangements, and the influence of rotation on performance. It has been claimed that load capacity is augmented by hydrodynamic action while gas flow remains substantially unchanged for a hydrostatic air bearing with rotation.

Similarly it has been stated that motion has a stabilizing effect, so that a stable stationary bearing will be stable at speed, but this seems unlikely to be generally true. The general questions of dynamic loading and stability of rotating hydrostatic bearings of different types offer scope for a great deal of research.

Mention has already been made of the potentialities of gas lubrication for very high temperature bearings. The only experimental confirmation is provided by the work of J. D. PIGOTT and E. F. MACKS¹⁹ on a stationary thrust bearing externally pressurized with air and tested at temperatures up to 540°C. There were considerable discrepancies between theory and practice, so that further investigation is needed.

Conclusion

This survey of gas lubrication shows that although oil will not suffer large-scale competition due to these new developments, gas bearings do show promise for many applications, in particular where loads are light, and speeds and temperatures are high. The simplicity of the self-acting air bearing, operating with the surrounding atmosphere as a bath of lubricant, is particularly attractive. In gas circulating machinery hydrodynamic gas bearings are an obvious choice, and in dynamometers and weighing devices hydrostatic air bearings offer advantages over similar bearings using oil. In ultracentrifuges and high speed machine tools, both types of air bearing have already shown their utility, while gas lubrication should be practicable and worthwhile in high speed electric motors. Future development may extend the use of gas bearings, particularly externally pressurized bearings, to plant for textile and food processing, where cleanliness is important, and also to steam and gas turbines, but the low load capacity and inability to deal with momentary overloads excludes the likelihood of use in heavier engineering plant.

There are more exotic possibilities in the field of transport. Several wheel-less vehicles, supported on air cushions produced by a ducted fan and jet propelled over land or water, have been publicly demonstrated recently, although only meagre technical information is available. The Curtiss-Wright *Air-Car*, the Ford *Levacar* and the Saunders Roe *SRN 1 Hovercraft* are all broadly equivalent to hydrostatic air bearings of low load capacity, about 0.1 lb per square inch of projected area, and very large power consumption, about 0.03 to 0.15 horsepower per pound of total weight. Only the Ford vehicle rides on an air cushion thin enough to be characteristic of film lubrication, i.e., thickness less than one thousandth of mean length, and conse-

quently it shows the best economy in power used. The Curtiss-Wright and Saunders Roe vehicles can hardly be said to involve true hydrostatic lubrication, for the load-carrying air cushion is generated by jet momentum effects and is relatively thick. Thus more power is required to lift a given weight, although the inwardly directed annular jet of the SRN 1 gives a good performance. The practical virtue of the thick air cushion is its greater tolerance for roughness of the surface over which the vehicle travels: the Ford vehicle in its present form is limited to specially prepared smooth surfaces. All the vehicles will become more efficient as the size is increased, for the cushion conditions will then approach those of film lubrication more closely.

The development of gas lubrication is based on long established and simple physical concepts but the practical evolution of suitable mechanisms and their detailed analysis for design calculation purposes have proved to be difficult. It is only the upsurge in technological development of the last two decades that has provided the stimulus for the recent advances in understanding and in application. The future in the field of gas lubrication is challenging both for the research worker and the designer.

The author would like to acknowledge that the writing of this article owes much to the collaboration in research on air lubrication with a former colleague, Mr John Kerr of the National Engineering Research Laboratory.

References

- ¹ HIRN, G. A. *Bull. Soc. Ind. Mulhouse* (1854) 26, 188
- ² — *Engng* (1885) 39, 118
- ³ ROBINSON, J. *ibid.* p. 130
- ⁴ TOWER, B. *Proc. Inst. mech. Engrs* (1883) 632; (1884) 29; (1885) 58
- ⁵ RAYLEIGH, *Engineer* (1884) 58, 167
- ⁶ REYNOLDS, O. *Phil. Trans. Roy. Soc.* (1886) 177, 157
- ⁷ KINGSBURY, A. *Mech. Engng* (1950) 72, 957
- ⁸ — *J. Amer. Soc. naval Engrs* (1897) 9, 267
- ⁹ HARRISON, W. J. *Trans. Camb. phil. Soc.* (1913) 22, 39
- ¹⁰ STONE, W. *Commonwealth Engr* (1921) 9, 115, 139
- ¹¹ DRESCHER, H. *Z. Verein deutsch. Ingenieure* (1953) 95, 1182
- ¹² BREWSTER, O. C. *Trans. Amer. Soc. mech. Engrs* (1953) 75, 518
- ¹³ FORD, G. W. K., HARRIS, D. M. and PANTALL, D. *Proc. Inst. mech. Engrs* (1957) 171, 93
- ¹⁴ For example, see *Proc. 1957 Conf. on Lubrication and Wear* London: Institution of Mechanical Engineers, 1957
- ¹⁵ FULLER, D. D. *Machine Design* (1947) 19, No. 6, 110; No. 7 117; No. 8 115; No. 9 127
- ¹⁶ GÜMBEL, L. and EVERLING, E. *Reibung und Schmierung in Maschinenbau* Berlin: Krayn, 1925
- ¹⁷ HENRIOT, E. and HUGUENARD, E. *C.R. Acad. Sci., Paris* (1925) 180, 1389
- ¹⁸ MONTGOMERY, A. G. and STERRY, F. *A.E.R.E. Report ED/R1671* (1956)
- ¹⁹ PIGOTT, J. D. and MACKS, E. F. *Lubric. Engng* (1954) 10, 29

AUTOMATIC CONTROL IN THE U.S.S.R.

R. H. TIZARD

This article records the impression of one member of a team* which visited the U.S.S.R. in May this year to study developments in automatic control. The visit, which was sponsored by the British Council and the Institution of Electrical Engineers in cooperation with the Academy of Sciences of the U.S.S.R., lasted a fortnight and included visits to establishments in Moscow, Leningrad and Kiev. A list of members of the team and the establishments visited are given in an appendix on page 358.

AS EVERY visitor to the Soviet Union comes away with an impression of the vastness of everything, from the country itself downwards, it is perhaps not surprising that I should come away with an impression of the vastness of effort being put into developments in automatic control. And since the launching of the first Sputnik no one will be surprised to learn of another technical field in which Russian work is not to be belittled. In fact advance in automatic control is extremely rapid, and the techniques and application should reach equality with the West in two or three years' time.

I was one of a team of six visiting the U.S.S.R. to study the subject, as an exchange visit organized jointly by the British Council and the Institution of Electrical Engineers, and by the U.S.S.R. Academy of Sciences. We spent two weeks in Russia during which we visited educational institutes, research and development establishments, and factories in Moscow, Leningrad and Kiev. Ours was a pioneering visit, indeed we were the first British visitors to at least one large research establishment, and an important purpose of it was to establish lines of communication and make personal contacts which would assist future visits, and open up a general interchange of knowledge and ideas in the field. In this we can claim to have been completely successful, largely due to the cooperation and friendliness of the Russian scientists and engineers. Everywhere we were received with great cordiality, and although the inevitable difficulties of translation sometimes made communication of knowledge and ideas ineffective, information was given to us fully and freely.

Technological Education

A good deal has been written and said about the extent of technological education in Russia and yet it took a visit to make me fully appreciate it; and this is my excuse for writing yet more. In Moscow University there are 22,000 students of whom 13,000

are science students, and these do not include those studying medicine or engineering, who are educated at technological institutes. The undergraduate course is for five years, and about ten per cent of students stay on for a further one or two years of postgraduate work. Although the normal age of entry is 18, quite a number are accepted up to the age of 35. As with all educational centres in Russia, it is very difficult to get an idea of the standard of the work because the examinations are mainly oral and there are no set papers. However we saw two of the electronics laboratories in the Physics Faculty, and discussed the students' experimental work with the professor, and this certainly gave the impression of a high standard. We were conducted round the University by a senior student who spoke excellent English. He had not been specially chosen, because there had been a slight hitch in the arrangements and the University had not expected us; he happened to be passing and volunteered for the job. We had a most interesting discussion with him, and it convinced us of the value to be obtained from the forthcoming exchange of university students which has been agreed between the British and Soviet governments. One of those sidelights which make personal contacts so worthwhile was the information that, had our student not spent the afternoon acting as our guide, he would have been painting his old car!

We also visited two technological institutes for the education of engineers. One of these, the Institute of Electric Power at Moscow—often less prosaically translated as the 'Institute of Energetics'—has 13,000 students who take a five and a half year course including 19 weeks of industrial practice, and it is concerned solely with training engineers for the power supply industry. In the whole of the U.S.S.R. 83,000 engineers graduated in 1957, and in the same year 816,000 graduate engineers were at work in industry.

Another strong general impression is of the way in which the Russians appreciate the value of

* The opinions expressed in this article are those of the author and not necessarily those of the team. An official report on the visit is to be published shortly

material incentives. This may surprise many people in this country. It starts at the earliest age with the use of the Pioneers. This is like all our youth organizations rolled into one, and it is important that there is only one. It provides evening entertainments and summer camps, and so on, and the right to join it is a special privilege granted to school-children only as a reward for good work. If a child is out, there is no alternative. The students at university or technical colleges are given a stipend to provide for board and lodging, and rises in this are again dependent on good work. In industry piece-work pay is carried to the logical limit; there is no basic rate and a man's pay is strictly proportional to his output. Rewards for inventions are at the disposal of factory managers, and are really handsome. And besides the opportunities to enter a technological career in Russia the incentives are there—it is the highest paid profession in the country as well as having the highest status, and the salary differentials are staggering by our standards. A professor may have a basic salary of 4000 or 5000 roubles a month; although difficulties over rates of exchange make comparisons very uncertain, this is roughly equivalent to £4000 to £5000 a year. He may easily earn a further 2000 roubles a month in consulting fees or as his personal share of sponsored research fees, and some earn a great deal more. The director of a research institute earns about 10,000 roubles a month basic, whilst any scientist who is honoured by being made an Academician, equivalent to our F.R.S., receives an honorarium of 5000 roubles a month for life, regardless of what else he may or may not earn and the maximum rate of income tax is 13 per cent.

Research and Development

Research and development in the Soviet Union is carried out in institutes which are responsible either to one of the Academies of Science or to an economic council; in the latter case they serve a particular field of industry. To a large extent, as one might expect, the theoretical work is done in the institutes of the academy, and the applied work in the industrial institutes. Nevertheless we found that the former were in close touch with practical problems, and the latter did plenty of fundamental work. We visited nine such institutes. Four of these were engaged solely in work on automatic control, and their total staff numbers varied from 300 to 700. The largest of them, the Ukrainian Institute of Automation at Kiev, has plans, fully approved, to expand to a total of 4000 in the next three or four years. We were shown the plans for new building which will start shortly; there will be a very large laboratory block, a workshop for small-scale

production of new instruments, and a housing estate for the staff.

The Russians are well known for their work on non-linear servo theory, and much of this together with other advanced work such as on self-optimizing systems is carried out at the Institute of Automation and Telemechanics, our principal host for the visit. The work of this Institute is known in this country through its journal *Avtomatika i Telemechanika*, which can be obtained in an English translation. Here we found analogue computers being extensively used both to back up theoretical investigations and to assist in the design of specific systems. There was considerable emphasis on problems of telemetering and telecontrol, and this had led to development of solid state devices to replace relays, and an interesting device for simulating complex switching circuits.

Electric Power and the Steel Industry

Everywhere we found an overwhelming priority given to two fields of application: electric power supply and steel. A large generator and network simulator is being used to study the problems of a grid system for Russia, being applied at present to investigate the seven-year plan project for lines connecting Moscow to Kubishev and the Urals, and to Stalingrad and Donbas. The alternators are represented by machines about 3 ft in diameter, designed to simulate the characteristics of full-scale machines; electronic simulators are used for the turbine characteristics and the voltage regulators, although the actual regulators and also protective systems may be connected to the model. The studies carried out so far have led to developments in the voltage regulators and in the methods of generator excitation.

The familiar problem of 'load despatching', so as to ensure the most economic operation of different stations feeding into a network, is probably even more important in Russia than in this country, because of the very wide variations in 'incremental costs'. Fuel used in power stations includes atomic energy, water power, coal, natural gas, and peat, and the problem is complicated by the fact that some stations are limited by the total amount of fuel available each day, rather than by their maximum generating capacity. To aid the load despatcher, an analogue computer has been developed and the first model is in operation at Kiev. In the analogue a unit for each station represents the function of incremental cost against load. There units are connected up, together with another unit which allows for network losses, in such a way that manual adjustment to the total load requirement auto-

matically causes the current through each unit to reach a value corresponding to the optimum load for the station; this is also automatically telemetered to the station. Further similar units enable the maximum fuel limits to be taken into account.

Work on boiler, turbine and generator control is being carried out, and a project has just started on computer control of boiler start-up. This was inspired by publication in this country of a similar proposal, and the Russians expect to have an experimental system in operation early this year.

In the steel industry control of all kinds of steel-making furnaces is being intensively studied. These studies appear to be taking a very logical form. To take as an example blast furnace control, the first step has been to gather data on the performance of a blast furnace, and to aid this automatic data logging equipment is being developed. This has included the construction of a very ingenious digital recording device which automatically transmits the weight of each charge delivered to the furnace. Such devices, used initially for experimental data collection, will eventually form part of the control system. The next step is to study theoretically and in practice the effects of a number of separate and independent controls, such as the control of temperature distribution from centre to outside of the furnace by varying the sequence of coke and ore charges; the stage has been reached where several such 'control loops' are being tried out in different steelworks. At this stage the necessity for certain new instruments is revealed, and as a consequence there is now under development, for example, an ultrasonic indicator of burden level. The final stage is the integration of the several control loops into an overall control system, probably using a computer, and here the Russians are at the stage of preliminary thinking of possible techniques.

The control of open-hearth and of electric-arc furnaces is following the same sort of pattern, in which advanced thinking is going hand-in-hand with solid practical experience. There are many interesting projects also in the control of steel-working processes. The characteristics of rolling mills are being studied theoretically (following, incidentally, original British work) and by electronic analogues, and there is a most interesting project for applying self-optimizing control to a 4-stand cold reduction mill.

By contrast with steel and power supply, automatic control in the chemical industry seems to be completely neglected, or at any rate we did not see any developments in this field. Amongst other work of interest were machine tool control, telecontrol of long gas pipelines, development of pneumatic instruments and of an automatic meteorological

station, and a number of things outside the strictly control field, such as digital computers.

Conclusions

Enough has been said to show that the Russian developments in automatic control are advanced. They have not yet caught up with the West, and indeed we saw very little which, in technical detail, would help our own developments. Our visits to factories were not adequate to allow us to judge the state of industrial application of control, but I have the impression that it is well behind ours. I believe also that much of the development work is based on faith in the newest techniques, which may pay big dividends in the long run, although more effort put into conventional techniques would give greater immediate benefits.

Behind it all is the immense effort being put into education and technology in general, and into automatic control in particular. The resources that we devote to automatic control are puny by comparison, even considered in proportion to the populations. As a result, progress in the Soviet Union is very rapid, and they cannot fail to catch us up in a very few years. Then the question is, will they surpass us? There are those who argue that the Russians are great imitators but poor originators. Even if this were true, is it any reason for being complacent?

Appendix

The team was led by Professor A. TUSTIN of the Imperial College of Science and Technology and the other members were Mr A. J. YOUNG (Imperial Chemical Industries Ltd), Mr S. S. CARLISLE (British Iron & Steel Research Association), Mr A. ASHBURY (English Electric Company Ltd) and Mr P. H. HAMMOND (National Physical Laboratory).

Visits were made to the following establishments:

Computer Centre of Academy of Sciences, Moscow
Experimental Laboratory of the Institute of Psychiatry of the Academy of Medical Sciences, Moscow
Moscow University Computing Centre
Moscow University Physics Faculty
Central Institute of Complex Automation, Moscow
Institute of Automation and Telemechanics, Moscow
Institute of Precision Mechanics and Computer Techniques, Moscow
Small Car Factory, Moscow
Nuclear Power Station, Obninsk, near Moscow
Institute of Electric Power, Moscow
Institute of Electromechanics, Leningrad
Electrotechnical Institute, Leningrad
Carburettor Factory, Kiev
Precision Instrument Factory, Kiev
Ukrainian Institute of Automation, Kiev
Department of Technical Science of the Ukrainian Academy of Sciences, Kiev
Ball-bearing Factory, Moscow
Central Laboratory of Automation, Moscow
Central Scientific Institute for Technology and Machine Building, Moscow

RESEARCH

The 1959 Waverley Gold Medal Essay Competition

The Scientific Advisory Board of RESEARCH have awarded the following prizes for the best essays in this competition.

The WAVERLEY GOLD MEDAL and First Prize

Dr ROBERT GROSS of the Department of Physics, The University of California, Berkeley, California, U.S.A. for

'A Study of Combustion in Supersonic Flow'

Second Prize

Mr W. E. BENNETT of the Atomic Weapons Research Establishment, Aldermaston, Berkshire for

'The Generation of Direct Current at High Potentials'

Special Prize for an Author under Thirty

Mr BRIAN CHALMERS of the Faculty of Engineering, University of Bristol

'Engineering Design by Digital Computers'

The prizes will be presented to the winners by Sir Frederick Handley Page, C.B.E., on Monday November 2nd at 5.30p.m. in the lecture hall of the Royal Society of Arts, John Adam St., W.C.2

The 1959 Essay Competition for Schools and Technical Colleges

This year TECHNOLOGY, published by The Times, joined RESEARCH in sponsoring this Competition and extending its scope to encourage entries from pupils at technical colleges and secondary modern schools. The Scientific Advisory Board of RESEARCH and the Editor of TECHNOLOGY have awarded the following prizes for the best entries in the three sections of the Competition.

SECTION I

First Prize to ROYDEN CLUTTERBUCK of the County Secondary School, Churchdown, Gloucestershire for

'An Examination of the Small Mammals of the Churchdown District, Gloucestershire'

Second Prize to CHARLES MOORE of Woodbridge School, Woodbridge, Suffolk for

'Observation of Sunspots'

SECTION II

First Prize to D. J. CLARE of Wymondham College, Wymondham, Norfolk for

'The Importance of Vegetative Reproduction in the Distribution of the Moss Bryum Argenteum'

Second Prize to MALCOLM LONGAIR of the Morgan Academy, Dundee for

'The Solubility of Calcium Sulphate and its Heat of Solution'

SECTION III

First Prize to DEIRDRE SPARROW of the South Hertfordshire College of Further Education, Barnet, Hertfordshire for

'Rediscovery of the "New" Metals'

Second Prize was not awarded

SURVEY

A Faster Boat for the Navy

In the past few months the inhabitants of Portsmouth have probably watched a small naval patrol boat with considerable interest. The *H.M.S. Brave Borderer* (see Figure 1) is the first of a new class of fast, manoeuvrable patrol boats of about 100 tons average displacement (96 ft length, 25 ft beam). The boat is powered by three Bristol Proteus marine gas turbines with a continuous rating of 2800 B.H.P.—maximum obtainable rating is 3500 B.H.P.—and skims over the surface of the water, twisting and turning at a speed of over 50 knots* leaving a twisting trail of 'heat haze' from the engines, and creating a disturbance that makes all other vessels in the vicinity feel somewhat uncomfortable.

These craft are designed for offensive operation in enemy coastal waters and for defensive role to protect our own coastal waters. It was stated that this is the first warship of any navy in the world to have a surface speed in excess of 50 knots. The very high degree of manoeuvrability required is achieved by an electric hydraulic steering gear in conjunction with three rudders. The helm can be put hard over at full speed. The boat is equipped with gyro compasses and Decca navigation aids. It carries four 21 inch torpedoes and one 40 mm Bofors or, alternatively, two torpedoes and two 40 mm Bofors.

The *Brave Borderer* has been built by Vosper Ltd for the Admiralty after an exhaustive design study carried out jointly by them in conjunction with the Admiralty. Considerable research was carried out with regard to water inclusion in the air intakes—at the press demonstration visitors were also warned not to drop small objects near the air filters—and the possible effect of salt spray on the blading material. The construction is basically of welded aluminium members as far as the load bearing structure is concerned. Planking is of mahogany, sheathed in fibre glass below the water line. The air intakes are also of moulded fibre glass construction. The shafting is in monel metal with small high speed propellers shrunk on to the shaft; this in a novel feature which resulted from the development work. The exhaust is through the upper portion of transom, protected by a shelf and adjustable hinged flaps to

* Admiralty security would not disclose the exact speed though this may well be considerably in excess of 50 knots.



Figure 1. Trials of *H.M.S. Brave Borderer* near Portsmouth. This boat which is powered by three Bristol Proteus marine gas turbines, each of 3500 maximum B.H.P., has a speed in excess of 50 knots. The boat was built by Vosper Ltd for the Admiralty.

protect the existing outlet in the event of one turbine being stopped. (It is possible to use the boat with one, two or three engines in operation.)

Cell Nucleus Symposium

A large United States contingent attended the conference on the cell nucleus† held at Cambridge on 31st August and 1st September under the auspices of the Faraday Society, and some of the most interesting papers came from Oak Ridge National Laboratory in Tennessee.

One paper that drew special attention was given by Dr BOLLUM of Oak Ridge, who showed that native deoxyribonucleic acid (D.N.A.) is not a suitable 'primer' for D.N.A. synthesis, but that after denaturation by suitable means, *e.g.* heat, the non-priming D.N.A. is converted to primer. It appears that the non-priming native D.N.A. is double stranded and the primer D.N.A. is single stranded, and the two can be distinguished by the correlation of primer activity with certain physical properties. All the enzymes so far known to polymerize deoxynucleoside triphosphates require primer D.N.A., and therefore, according to Dr Bollum, the conversion of native D.N.A. to primer is also required. The natural mechanism for this is not yet known.

Of the five sessions that constituted the conference, the one on the structure and functions of the nucleolus aroused considerable interest, papers being given on the role of the nucleolus in mitosis, gastrulation, ribonucleic acid (R.N.A.) and protein synthesis, *etc.* The section on nuclear enzymes, with papers by

† This Symposium will be published in full early in 1960 by Butterworths Scientific Publications

Dr MIRSKY and Dr ALLFREY of the Rockefeller Institute and Dr SIEBERT of the Johannes Gutenberg University, was also very stimulating.

Some beautiful electron micro-photographs of chromosomes, showing them to consist, as he put it, of 'a hierarchy of pairs of pairs of pairs of strands', tidily coiled together, were shown by Dr KAUFMANN of the Carnegie Institution of Washington, and the conference closed with a forward-looking paper by Dr PHILPOT of the Radiobiological Research Unit, A.E.R.E. on the progress towards the culture of chromosomes and nuclei *in vitro*.

Silicon 'Stepping' Transistor

The production of a new four-terminal $p-n-p-n$ transistor has been announced by Bell Telephone Laboratories. The transistor, which is illustrated in Figure 2, performs the same functions as a 'gas stepping valve'; it acts as a pulse controlled on-off switch and will, in many cases, replace more elaborate circuitry.

The gas stepping valve uses the bistable voltage current characteristic of a gas discharge for its operation while the transistor utilizes a $p-n-p-n$ arrangement as the bistable element. This design results in a bistable voltage-current characteristic between a single common electrode and a set of

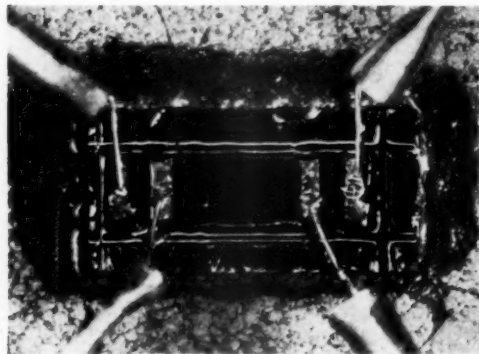


Figure 2. Stepping transistor element mounted on a gold plated Kovar header is shown magnified to about 100 times the actual size. The leads are thermocompression bonded to gold-silver alloy contacts (reduced to $\frac{1}{3}$ of original size by reproduction).

multiple electrodes. Nonsymmetrical geometry is employed to obtain a unidirectional transfer of voltage.

The transistors will operate within the range 1 to 100 milliamps with supply voltages of 10 to 100 volts and they have been operated at speeds of

up to one million pulses per second. The transistors are designed for use as a basic element in building up certain logic circuits in digital computers such as for counting and de-coding. One transistor can be used to drive two others and thus provide more versatile de-coders. The stepping transistors can be made either of interconnected elements or fabricated on a single piece of silicon.

Polymer

A new quarterly journal, *Polymer*, is to be published by Butterworths Scientific Publications beginning in January, 1960. The journal will contain both full length original papers on pure and applied polymer research and short advance publications. Members of the Editorial Board, which consists of Dr C. H. BAMFORD of Courtaulds Ltd, Professor C. E. H. BAWN of the University of Liverpool, Professor GEOFFREY GEE of the University of Manchester and Dr ROWLAND HILL of L.C.I. Fibres Division, will advise on the papers to be included. The price per volume is £5 and each volume will contain about 500 pages.

Cosmic Radiation and Tuberculosis

The Chinese journal *Science Record*† is published monthly and contains short papers on a wide range of scientific subjects. It covers mathematics, chemistry, physics, biophysics and geology. The list of contents is usually interesting and the January 1959 issue has one item that may have importance in the field of medical research. It deals with investigations by S. G. Ong, which have obviously covered several years, on the effect of cosmic radiation on tubercle bacilli.

Ong has shown that when tubercle bacilli were exposed at the Jungfrauoch (3457 metres) with or without screening there was a significant difference in virulence. The mean time of survival (in days) when injected into mice was taken as a measure of their virulence.

Direct cosmic radiation	116 days
Under 2 cm of lead	79.9 days
Under 30 m of rock	20.5 days
Under 10 cm of lead	bacilli killed

Similar experiments were carried out at Peking at 63 metres above sea level, but the effect here was much less marked. The author also states that preliminary experiments have shown that mice injected with tubercle bacilli and exposed to cosmic radiation under a protective lead shield also showed a strong immunization effect against subsequent injections with virulent tubercle bacilli.

† In English with a few papers in other European languages. Published by Science Press, Peking, China.

SIR ALFRED EGERTON, F.R.S.



SIR ALFRED EGERTON died on Monday, September 7th. He had been a member of the Scientific Advisory Board since the first meeting on November 25th, 1946.

Egerton was a scientist and technologist of eminence in his chosen field, combustion. He had trained under Ramsay at University College, London, Gutz at Nancy and Nernst in Berlin. Before going to Berlin he had been an instructor in chemistry at the Royal Military Academy at Woolwich and while there he worked in Sir William Ramsay's private laboratory in Regent's Park, being the last of Ramsay's research assistants. His interest in combustion and particularly in the problem of 'knock' dates from 1924 when he was at Oxford. He showed that antiknock materials act as inhibitors of combustion and related this to the new conceptions of chain reactions and the importance of the peroxide radicals in combustion. He held the Chair of Chemical Technology at Imperial College for sixteen years, from 1936 to 1952. During this time both his reputation and that of his department grew to its present international proportions. And during much of this time, from 1938 to 1948, he was Secretary of the Royal Society for the Physical Sciences. He was the Editor of *Fuel* and the Editor-in-Chief of *Combustion and Flame*.

It is as a friend of and a friend to *Research* that we remember him. Egerton was a very remarkable man, not the least in his wide humane and artistic interests. Indeed the most effective concealment of his remarkable qualities was presented by Egerton himself. A scientific career was not a common project among Etonians born in the eighties and a career devoted to research in chemical technology was even less likely. His kindness and consideration were infinite. His resolution of character, concealed by a diffident charm of manner, could be shattering but was never meant to be. It was simply that Egerton thought in simple direct terms and expressed what was obvious to him. Simplicity is often shattering to those who have become enmeshed in complication. He will be missed by many deliberative bodies and the Scientific Advisory Board of *Research* will be very much the poorer to have lost his advice, his insight, his humanity, his charm. We would like to think he enjoyed the meetings, as we did because of his company.

BOOK REVIEWS

Elements of Physical Metallurgy

A. G. GUY

(xvi + 528 pp; 9½ in. by 6 in.)

Cambridge, Mass.: Addison-Wesley. \$9.50

FORTY years ago 'physical metallurgy' was concerned mainly with the examination of industrial metals and alloys by the methods of microscopical metallography, thermal analysis, dilatometry, and mechanical testing. In the ensuing years the meaning of the term has been enormously extended, and now includes an immense field ranging from the electron theory of solids to the problems of industrial metallurgy. There is thus a great need for a book which will enable the university student to survey the subject as a whole, and to learn how the fundamental work is related to practical problems.

The present book is an attempt to supply this need, but is only partly successful. Any student who works through the book systematically will learn a great deal, and chapter 8 provides an elementary introduction to elasticity which may be of real value to those who find difficulty with more advanced books. In many chapters, however, the writing lacks conciseness and clarity and, apart from some definite mistakes, there are many places where a definition is incomplete, or a wrong meaning is easily gained. To the British reader, the value of the book is greatly reduced by the fact that nearly all temperatures are given in degrees Fahrenheit.

W. HUME-ROTHENY

Radio Studies of the Universe

R. D. DAVIES and H. P. PALMER

(viii + 200 pp; 8½ in. by 5½ in.)

London: Routledge & Kegan Paul. 25s

ACCORDING to its foreword this book should be understood by the intelligent non-scientist and the style gives the impression that the authors had this in mind. The main part (chapters IV to IX) gives a concise and clear account of the contemporary contributions to our knowledge of the structure of the universe from radio-astronomy but is likely to be heavy going for the layman. The intelligent undergraduate in physical science should certainly find his interest stimulated and the optical astronomer who wishes to widen his background in this direction will find here an excellent introduction. In fact it may have been better if the book had been consciously directed to groups such as these, since it would have avoided a certain unevenness in presentation, particularly where the authors have attempted explanations in simple terms of difficult physical concepts. For instance it is quite misleading to describe the scintillation of radio stars as due to a 'shadow pattern'—even with the qualification 'more accurately described as the diffraction pattern'—especially when the accompanying diagram (Figure 9.3) depicts a diffraction pattern resembling in detail the shadow of the clouds of ionization producing it.

The book is least successful in dealing with the techniques used in radioastronomy; it is especially here that

the tendency to oversimplify does not necessarily make the subject more intelligible to non-specialists. Few of these will be aware of what is meant by the 'noise generated in a matched input resistance at room temperature', while they will get a quite incorrect impression from the statement that the dimensions of aerials are chosen partly to render them less likely to pick up signals on wavelengths other than the desired one. A number of other blemishes could have been excluded by more care in preparation. They include a rather odd looking radio noise waveform in Figure 3.4, an interferometer record in Figure 3.6b of the type produced by Ryle's phase switched adaptation of the Michelson interferometer whereas the instrument described and depicted in Figure 3.6a is a conventional one, and an incorrect value for the earth's orbital velocity in Figure 8.6.

The fourteen excellent plates include two of the large radiotelescope at Jodrell Bank and the wise inclusion of a glossary may help the uninitiated. I can recommend the book as a useful contribution to the growing volume of literature on the subject.

T. R. K.

Experimental Techniques in Low Temperature Physics

G. KENDALL WHITE

(viii + 828 pp; 8½ in. by 5½ in.)

London: Oxford University Press. 45s

Now that liquid helium and helium liquefiers can be bought, it is no longer necessary to belong to an experienced university department in order to conduct experiments at temperatures below 60°K. Engineers and chemists are joining the physicists in finding good reasons for entering this field, and it is to those with this intent that this book is addressed, and to whom it can be recommended. Dr WHITE provides a sustained account of the experimental and design procedures required for the successful use and handling of liquefied gases, with a copious supply of references and illustrative calculations. There are chapters on the design of cryostats, on the storage and transfer of liquefied gases, on adiabatic demagnetization for temperatures below 1°K and on temperature control and measurement. The last two chapters contain collected data on the heat capacities, expansivities, and electrical and thermal resistivities of materials in common use.

D. H. MARTIN

From Microphone to Ear (2nd Edn)

G. SLOT

(ix + 258 pp; 8½ in. by 5½ in.)

London: Cleaver-Hume Press (Philips' Technical Library). 21s

From Microphone to Ear is intended primarily for readers with little or no technical training. The author has therefore avoided mathematics as much as possible and relied on a large number of diagrams and simple, logical explanations. Consequently, any reader interested in

understanding the principles and practices of modern sound recording should be able to read this book with enjoyment. The main emphasis is on commercial disc recordings, and the first section of the book contains a brief account of the manufacture of these records. More detailed chapters describe both the mechanical construction and performance of pick-ups, microphones, amplifiers, and loudspeakers. The chapters on the needle, the record and record players contain an interesting account of some of the distortions which are inevitable in reproduction from a spiral groove cut in a resilient material. The advice on the care of both needles and records will be both valuable and reassuring to owners of record collections.

This edition contains a new chapter on stereophonic reproduction. Following a short description of basic principles, the chapter describes the two-channel gramophone records now commercially available and the particular problems which arise when two separate signals are carried by one spiral groove.

The book contains few mistakes, although there are several grammatical and spelling errors which must be attributed to the printing of an English language book in Holland. One part of the book falls below the high standard of clarity of the remainder. This is the short introductory section which explains the only mathematical concept used throughout the book—the decibel. Many non-technical readers will not understand the nature of a logarithm of a ratio and therefore will not realize that the decibel is a relative unit, with a zero level chosen for convenience. There is another difficult passage on page 55 where the author has chosen to emphasize the importance of the shape of a gramophone pick-up arm, rather than its effective length and the position of the pivot, when discussing methods of reducing tracking distortion.

H.N.D.

Magnetic Sound Recording

D. A. SNEL

(xii + 217 pp; 9 in. by 6 in.)

London: Cleaver-Hume Press (Philips' Technical Library). 25s

Magnetic Sound Recording is also in the Philips' Technical Library series, but when compared with *From Microphone to Ear* is most disappointing. It is not clear for whom the author intended his book, and the technical level varies considerably. Some sections are extremely over simplified, others are difficult to understand. The introduction to the decibel is confused by mention of the 'phon' and concludes with a sentence which includes the phrase '... the ratio of loud to soft is in effect negative ...'.

The author's introduction to the theory of magnetism relies too much on sections of mathematics which are inadequately explained. These calculations are expressed in MKS units without any explanation of this system of units. Many readers will be puzzled by such units as the newton and the weber, particularly as a table on page 33 contains values in gauss.

The whole book suffers from a disjointedness and vagueness which seems partly due to inadequate re-editing for the English market. References to the value of the earth's magnetic field in Holland, to German radio receivers, to the BBC as an American company, and to the South Magnetic Pole in Canada are typical of this. Although the book will be of interest to many readers wishing to learn more about tape recorders, it cannot be recommended in its present form. A future edition could be greatly improved by re-editing and by increasing the technical content at the expense of the last 28 pages of text which contain little more than lists of possible uses of tape recorders.

H.N.D.

Modern Aspects of Electrochemistry No. 2

J. O'M. BOCKRIS

(ix + 416 pp; 8½ in. by 5½ in.)

London: Butterworths Scientific Publications. 75s

LIKE the first with this title, which appeared in 1954 under the joint editorship of J. O'M. BOCKRIS and B. E. CONWAY, this book forms part of the Modern Aspects Series of Chemistry under the general editorship of F. C. TOMPKINS, and is similar in nature, in that it does not all come from the same pen, but differs from it in the specialized topics dealt with in the subject matter. It maintains in almost every respect the high standard and production set by earlier volumes in the series.

The present volume is divided into five chapters or sections, each of which constitutes an up to date and extensive review of an aspect of electrochemistry which is either new or has come to the forefront in recent years. As the various sections have been written by active workers and leading theoreticians in their respective fields, the whole book bears the stamp of authority and informed opinion at a high level.

Section 1 (86 pages) is entitled *The Present State of Electrolytic Solutions* and is by H. FALKENHAGEN and G. KELBG. This is a very compact and masterly treatment of the subject, well illustrated and with an emphasis on the mathematical aspect. Recent theory of concentrated solutions and interionic attraction receive a very full treatment. Section 2 (73 pages), *Physical Chemistry of Ion-exchange Resins*, by J. A. KITCHENER, is a timely review of a new and practically important field, the electrochemical basis of which is beginning to take an ordered shape. Section 3 (102 pages), *Molten Electrolytes*, by H. BLOOM and J. O'M. BOCKRIS, condenses the results of an enormous mass of recent research—much of it instigated by the demands of atomic technology—into a readily digestible form. Section 4 (81 pages), *The Anodic Behaviour of Metals*, by T. P. HOAR, is concerned essentially with the mechanism of anodic processes at electrified interfaces, a phenomenon which has an important bearing on the subject of corrosion. Section 5 (65 pages), *Electrochemistry of the Semiconductor-Electrolyte Interface* by M. GREEN, concludes the book with the first review to be published on a hitherto much-neglected topic, one which is at long last beginning to receive some attention.

The style is generally good, although the English in Section I is sometimes rather odd and bears a distinctly continental flavour. But this is a minor criticism, as the meaning is almost everywhere perfectly clear. The book is abundantly provided with clear line drawings and diagrams, as well as a useful subject index. The references listed at the end of each section are copious, varying approximately from 100 to 350. This book should be found in every science library, and is warmly recommended to the specialist.

L. H. D.

The Chemistry of Natural Products: II Mono- and Sesquiterpenoids; III The Higher Terpenoids

P. DE MAYO

(vii + 320 pp; vii + 239 pp; 9½ in. by 6 in.)

New York and London: Interscience. \$7.50 and \$6.00; 52s and 42s

A STUDY of the terpenes reveals a fascinating field of chemistry. It also provides examples of many methods of degradation and synthesis, illustrates the applications of physical measurements, and leads the student to many of the theories of reaction mechanisms. As the author states in his preface, 'correctly studied, terpenoids present almost a microcosm of organic chemistry'.

Until the appearance of these volumes information of a reasonably detailed and up to date nature about the various groups of terpenes, and more especially about those groups of sesqui-, di- and triterpenes which have been so intensively studied in the last ten years, was to be found only in various monographs or in reference books such as Simonsen's *The Terpenes*. The author is to be congratulated for providing in these two volumes a comprehensive and critical survey up to 1956 suitable for both the honours student and lecturer, the research student and supervisor.

The first of the two volumes (Vol. II of the series) which is edited by K. W. BENTLEY commences with a brief survey of the sources and isolation of terpenoids. After a passing reference to the general methods of degradation, there follows a most useful survey of the applications of ultra-violet, infra-red, and nuclear magnetic resonance spectroscopy and of molecular rotation data. Next come chapters on the acyclic, mono- and bicyclic terpenoids. In these chapters succinct accounts of conformational analysis and Wagner-Meerwein rearrangements are to be found. These rearrangements are also discussed further in a separate chapter along with the theories of classical and non-classical carbonium ions. The volume ends with two chapters on sesquiterpenoids.

The second volume (Vol. III of the series) covers the diterpenoids, the tetracyclic triterpenoids, the α - and β -amyrin and lupeol group of pentacyclic triterpenoids and finally other pentacyclic triterpenoids. The volume ends with a review of the biosynthesis of terpenes. The

recent work in this field has shown that steroids are intimately related to the triterpenoids and has also provided an explanation of the origin of the 'isoprene' unit. Unfortunately the review gives an incomplete picture of the *current* position since during the time taken to publish these volumes many advances have been made in the study of the biosynthesis.

The two volumes are not free of mistakes. The most glaring concerns the very first structure given in the volume on monoterpenoids. This structure is supposed to represent isoprene, the five-carbon unit of which has played such a big role in terpene chemistry, but it represents, in fact, 2-methylbut-2-ene. Another obvious slip concerns the prefaces. The same preface has been used for both volumes without appropriate modification. However such blemishes do not detract from the fundamental value of these volumes; a purchaser of these volumes will get full value for his money.

T. G. HALSALL

River Pollution I: Chemical Analysis

L. KLEIN

(ix + 206 pp; 8½ in. by 5½ in.)

London: Butterworths Scientific Publications; New York: Academic Press. 30s; \$3.50

DR KLEIN's latest book *River Pollution I: Chemical Analysis* is as he says in his preface, a revision of chapters 9 and 10 of his earlier and more comprehensive work *Aspects of River Pollution*. These chapters dealt with the application and interpretation of chemical methods of analysis in the study of river pollution and in the treatment of polluting liquids. In the new book the same ground is covered using, to a large extent, the original test, supplemented where appropriate by sections summarizing work published since 1956. New material includes many of the methods put forward in *Recommended Methods for the Analysis of Trade Effluents* by the Joint Committee of the Association of British Chemical Manufacturers and the Society for Analytical Chemistry, and a fair number of other new methods with corresponding extra references to original papers.

The analytical methods dealt with include determination of dissolved oxygen, the principal oxygen demand tests, and estimation of alkalinity and acidity, metallic contaminants, compounds of nitrogen and sulphur, and a variety of compounds of importance in work on pollution. In the last chapter the reasons for making the tests and the meaning of the results when obtained are discussed with some indication of precautions which may have to be taken in certain circumstances.

The value of the work might, perhaps, have been enhanced if more critical comments had been included in the descriptive chapters to assist the analyst in selecting from alternative methods the one most suitable to a given set of conditions.

E. J. PETTET

NOTICE TO CONTRIBUTORS: All communications intended for the Editorial Department should be addressed to The Editor, *Research*, 4 and 5 Bell Yard, London, W.C.2. Contributions sent on approval will be carefully considered by the Editor; but no responsibility whatsoever can be accepted for manuscripts, although every effort will be made to return them if a stamped addressed envelope is enclosed for that purpose. The copyright of all contributions for which a fee is paid shall belong to Butterworths Publications Ltd, together with the right of republication. The Editor reserves the usual right to make alterations in the text of articles accepted for publication. The author will receive galley proofs and if alterations have been made every effort will be made to obtain the author's prior agreement. Unless special arrangements are made, contributions are only received and considered on these conditions.

re
so
'
te
ne
ve

st
ne
ed
as
it
us
en
n.
a-
se
L

w

al
rs
rk
ne
of
it-
ne
al
ns
al
n-
py
sh
al
ds

n-
nd
ic
r,
on
ng
ed
ch

en
ed
ng
a
ET
h,
er
se.
n.
nd
le,